



Recovering biological electron transfer reaction parameters from multiple protein film voltammetric techniques informed by Bayesian inference



Henry O. Lloyd-Laney^a, Nicholas D.J. Yates^b, Martin J. Robinson^a, Alice R. Hewson^b, Jessie Branch^c, Glyn R. Hemsworth^c, Alan M. Bond^d, Alison Parkin^{b,*}, David J. Gavaghan^a

^a Department of Computer Science, University of Oxford, Wolfson Building, Parks Road, Oxford OX1 3QD United Kingdom

^b Department of Chemistry, University of York, Heslington, York YO10 5DD, United Kingdom

^c Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, United Kingdom

^d School of Chemistry, Monash University, Clayton, Vic. 3800 Australia

ARTICLE INFO

Keywords:

Voltammetry

Inference

Bio-electrochemistry

ABSTRACT

Deciphering the mechanism, kinetics and energetics of biological electron-transfer reactions requires a robust, rapid and reproducible protein-film voltammetry information recovery process. Here we describe a semi-automated computational approach for inferring the chemical reaction parameters for a simple protein system, a bacterial cytochrome domain from *Cellvibrio japonicus* that displays reversible one-electron $\text{Fe}^{2+/3+}$ redox chemistry. Despite the relative simplicity of the experimental system, developing a robust data analysis approach to find the global optimum in 13-dimensional parameter space is a challenging task because the Faradaic-to-background current ratio in such experiments is often low. We describe how a multiple-technique approach, whereby data from three voltammetry techniques (direct-current, pure sinusoidal and Fourier transform alternating current voltammetry) is combined, ultimately enables the automatic extraction of both (i) quantitative “best-fit” redox reaction parameter point values that are robust across multiple experiments performed on different protein-electrode films, and (ii) a statistical description of parameter correlation relationships, along with uncertainty in the individual parameter values, obtained using Bayesian inference. It is the latter achievement which is particularly important as it represents a method for visualising the possible limitations in the mathematical model of the experimental system. Our multi-voltammetry analysis approach enables such powerful insight because of the complementarity between the information content, simulation-speed and parameter sensitivity of the current–time data generated by the different techniques, illustrating the value of adding purely sinusoidal voltammetry to the bioelectrochemistry measurement toolkit.

1. Introduction

Electron-transfer reactions underpin many of the fundamental chemical processes that are required for the existence of life, such as the splitting of water, the fixation of nitrogen, and carbon capture. Such catalytic redox processes are of significant biotechnological interest because they are achieved at ambient temperatures and pressures using commonly available metals at the catalytic active 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 [1]. In voltammetry, a time-varying potential is used to drive electron-transfer reactions, so that the recorded experimental current–time data directly reports on the reac-

tions that take place. In this paper, we seek to define a voltammetric modelling framework that incorporates a complementary set of experimental measurement techniques and computational analysis tools, with the aim of extracting as much accurate information as possible about biological electron-transfer reactions. We demonstrate how the inclusion of purely sinusoidal voltammetry into the suite of protein film electrochemistry techniques enables Bayesian statistical analysis to be used as a quantitative means of visualising uncertainty in the voltammetric modelling.

In voltammetric modelling, the aim is to reconstruct the underpinning reaction process that generated the experimentally measured data [5]. For protein film voltammetry, the resulting mathematical model takes the form of a system of non-linear ordinary differential equations. Embedded within this system are the key parameters of interest

* Corresponding author.

E-mail addresses: alison.parkin@york.ac.uk (A. Parkin), david.gavaghan@dtc.ox.ac.uk (D.J. Gavaghan).

<https://doi.org/10.1016/j.jelechem.2023.117264>

Received 15 December 2022; Received in revised form 20 January 2023; Accepted 8 February 2023

Available online 20 February 2023

1572-6657/© 2023 The Author(s). Published by Elsevier B.V.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

that govern the biological redox reaction — that is, the electron transfer mechanism, reaction rates, and electrochemical potentials that directly report on the thermodynamics of the reaction [6,7]. For the one-electron reaction studied here, these parameters are detailed below, with their mathematical relationships detailed in the SI. Using a single set of input parameter values, it is possible to simulate (typically computationally) a current–time–potential output — 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 — known as the inverse or parameter inference problem [5,8]. This involves finding an optimal set of parameters that minimises a distance (or “objective”) function between the experimental and simulated data.

In protein film voltammetry, solving the inverse problem is challenging since the experimentally-measured current–potential–time data result not only from the Faradaic signal of interest, but also a typically very large “background” current, arising from charging at the electrode-solution interface, compounded by the effects of uncompensated resistance [9–13]. To fully compensate for such non-Faradaic processes, terms modelling capacitance and resistance are incorporated into the mathematical model, and these parameters must also be estimated as part of the inverse/inference process. This typically results in a very high-dimensional problem (in this study the inference problem is in 13 dimensions). 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 [14], and can be additionally complicated by spurious Faradaic processes, such as from quinones [15]. In the field of biological electrochemistry there is also a great deal of literature, including detailed review papers [3,4], exploring the difficulty in establishing the correct mathematical voltammetric model; there are challenges in (a) incorporating kinetic and/or thermodynamic parameter dispersion [16–18], (b) distinguishing between different catalytic mechanisms and inactivation reactions for redox enzymes [19–21], and (c) substrate transport [22,23]. As has been demonstrated previously [17,18], several of the parameters in a mathematical model of protein/enzyme film electrochemistry have similar effects over the potential window of interest and finding an optimal solution to the inverse problem is therefore made very challenging by the presence of multiple local minima values for the objective function, located in disparate parts of the overall parameter space. This study describes a new approach to overcome these problems; by analysing a complementary set of voltammetric measurements we aim to overcome low Faradaic current-to-background limitations, rapidly extract a set of “best-fit” redox-reaction parameters, and also use statistical tools to probe model uncertainty.

In previous work on surface-immobilised ferrocene, we demonstrated the advantages of combining Purely Sinusoidal Voltammetry (PSV) and ramped Fourier Transform Alternating Current Voltammetry (r-FTACV) when using voltammetric modelling to solve the inverse problem for film voltammetry data, demonstrating the substantial increase in speed of analysis that is possible when assessing PSV data relative to r-FTACV [18]. This decrease in simulation time is driven by the reduced number of sinusoids required for a complete experiment — 30 for PSV, vs. 412 for r-FTACV in the experiments analysed in this paper. The time reduction is important because the slow rate of r-FTACV analysis had previously made it practically unfeasible to model high frequency experiments that enable quantitative determination of rapid rates of electron transfer [24]. PSV and r-FTACV are particularly useful when analysing protein film voltammetry data as they show high sensitivity to Faradaic parameters of interest. The sensitivity of PSV and r-FTACV predominantly arises from the non-linearity of the current response to a sinusoidal potential with large amplitude — when the current is Fourier transformed, harmonics are observed at multiples of the frequency of the input potential [1,18,25]. As the background current is less highly non-linear, appropriate harmonic

selection can yield a signal that reports exclusively on the Faradaic process of interest. The extension of our work to a protein system additionally required the incorporation of direct current voltammetry (DCV) measurements into our experiment-analysis protocol. We use DCV to estimate electron transfer rates and protein-electrode coverage values in a manner analogous to the work pioneered by Armstrong and co-workers [11,12], and building on historic voltammetric theory by Laviron [26].

There are alternative electrochemical measurement methods upon which we could have based our technique development for the rapid inference of “best-fit” model parameter point values: square wave voltammetry (SWV) also amplifies the Faradaic-to-background current sensitivity of a voltammetric measurement albeit via a different mechanism to PSV and r-FTACV [27], while electrochemical impedance spectroscopy (EIS) is widely used to probe the capacitance–resistance model of an electrochemical system [28]. Indeed, there are numerous bioelectrochemistry studies which have made powerful use of DCV [10,9,4], SWV [29,30] and EIS [31,32]. We and others have previously shown how r-FTACV alone can be applied to analyse protein electron transfer [33,1,24]. In this study, the reason we use PSV is because extending the PSV fitting procedure to include Bayesian analysis is relatively simple, meaning that we can generate inferred 1D parameter histograms to obtain confidence limits for parameter values, and also 2D histograms to show parameter correlation relationships (this is explored in more detail in the SI in Figs. S1 and S2). This is the first time such an exploration of parameter space has been achieved in a bioelectrochemical context, although such statistical analysis has been applied in EIS [34], small-molecule voltammetry [35] and other domains of chemistry [36].

The experimental system characterised in this study is a simple protein from *Cellvibrio japonicus*, referred to as CjX183 [37]. This is a type-c cytochrome domain of Cbp2D, a probable activating partner for a cellulose degrading enzyme known as a lytic polysaccharide monooxygenase (LPMO) [37]. LPMOs facilitate industrial biofuel production from biomass [38]. It has recently been shown that CjX183 can transfer electrons to LPMOs via the reversible one-electron redox reactivity of the heme centre [37]. DCV measurements of the CjX183 $\text{Fe}^{3+/2+}$ redox chemistry were obtained previously via adsorption of the protein onto a pyrolytic graphite edge electrode [37], and here we use the same immobilisation approach to probe this redox process using more complex potential inputs, to enable a correspondingly more powerful interrogation of the Fe-centered redox chemistry.

2. Experimental methods

The CjX183 protein was purified as per the established protocol in work by Hemsforth [37] 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. Additional details can be found in the SI.

A custom-built electrochemical cell was set-up exactly as previously described for studies of wild type CjX183 [37], inside an anaerobic chamber (< 40 ppm oxygen). A 10 μL aliquot of protein was pipetted onto the working pyrolytic graphite electrode tip and left to adsorb until a film has formed. The cell was maintained at 5 °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 reference electrode (Ref) was a saturated calomel electrode filled with aqueous 3 M KCl solution. Calibration experiments with ferricyanide gave a value of $E(\text{Ref}^+ = E(\text{SHE}) + 0.239 \text{ V}$. Within this paper, all potentials are reported in V versus Ref. The DCV experiments were carried out initially after protein adsorption and then after r-FTACV and PSV experiments. These 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 s 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} .

The r-FTACV experiments were conducted using a custom potentiostat and the instrument was controlled using custom software [25]. Each r-FTACV experiment commenced with a 5 s 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 known frequency close to 9 Hz and amplitude of 150 mV [25].

The PSV experiments were conducted using the same potentiostat and software as for r-FTACV, using a 5 s 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 4.71 rads for 26.8 s (equivalent to 268 oscillations). As in previous work, the PSV current was truncated to the first 30 oscillations, to reduce simulation times [18].

Forwards problem simulation code for the model derived in the SI was written using a combination of C++ and Python, and all inference was performed using the PINTS repository [39]. Code for reproduction of results from the paper, along with experimental data, can be found on Github [https://github.com/HOLL95/Cytochrome_paper_results].

3. Results and Discussion

3.1. Approach for parameter recovery

Sequential application of DCV, r-FTACV and PSV to an electrode functionalised with CjX183 was repeated for three different functionalised electrodes, generating three experimental datasets, referred to hereafter as experiments 1–3. For each experimental run, the electrode was also subjected to the three voltammetry techniques prior to functionalisation with CjX183 (i.e. a “blank” electrode). The potential inputs that define the three techniques are shown in Figs. 1 A-C respectively. The resulting total current obtained during experiment 3, in the presence and absence of CjX183 (in blue and red respectively) is shown for the three techniques in Fig. 1D-F. The harmonics of the r-FTACV and PSV current responses can be observed in the Fourier domain in Figs. 1G and H, and Figs. 1J and K show those same harmonics in the time domain. As reported previously [40], in experiments without protein there is still a non-linear component to the current, resulting in significant background current contributions to harmonics 1–3. In the mathematical model, we incorporate the non-linear pseudo-capacitive contributions as a third-order polynomial. In experiment 3, multiple DCV experiments were conducted at different scan rates, and the oxidative/reductive peak position was extracted. These are presented in Fig. 1I, in the form of a trumpet plot.

The starting point for the data analysis was the derivation of the forwards problem simulation model of the single-electron $\text{Fe}^{2+/3+}$ redox process undertaken by CjX183, and this is described in the SI. The model assumes Butler–Volmer kinetics and incorporates thermodynamic dispersion, assuming a normal distribution of E^0 values, defined by a mean μ and a standard deviation σ [17,41]. In previous work, it was determined that explicit modelling of thermodynamic dispersion had to be included when fitting film r-FTACV voltammetry data [41,17,18]. This has been attributed to the range of protein orientations that can be achieved on the rough graphite electrode surface [41,17,18]. Other relevant Faradaic parameters include k^0 , the rate at which the redox reaction occurs at E^0 , α the symmetry factor, which is a measure of the relative ease with which the transition state is formed and Γ , the surface coverage of electroactive species on the elec-

trode. Background current parameters include C_{dl} , the magnitude of the background current arising from linear double-layer capacitance effects, and C_{dlEX} , which model non-linear capacitance effects (X is the order number, as discussed in the SI). R_u corresponds to the uncompensated resistance, ω the frequency of the input sinusoid (for PSV and r-FTACV), η the phase of the Faradaic current and correspondingly $C_{dl}\eta$ the phase of the capacitive current.

Accurate parameters inferred from a voltammetry dataset should correctly predict the current response of the interrogated film to any potential input. This was demonstrated in previous work on surface-confined ferrocene, when fitting PSV total current a single best-fit parameter vector could be obtained that also predicted r-FTACV harmonic data collected for the same electrode film. As detailed in the SI and Fig. S3, obtaining a similarly good fit for total PSV current of CjX183 was not possible, this is ascribed to the much lower Faradaic-to-background current ratios arising from voltammetry of a biological, rather than a small molecule, system. This decrease in Faradaic current arises due to the much larger molecular footprint of a protein compared to a small molecule, resulting in a smaller electrode film coverage density, and more dominant background currents. Subsequent attempts to solve the inverse problem for CjX183 initially involved unconstrained fitting of a filtered portion of the PSV total current, obtained by selecting harmonics 4 to 10 in the Fourier spectrum. This did not yield parameters that could also predict an r-FTACV experiment, details of which can again be found in the SI and Fig. S4. In response to these challenges, an iterative fitting procedure was developed.

3.1.1. The iterative parameter inference loop

The core loop of this iterative parameter recovery process was to fit to PSV harmonic data in the frequency domain (as opposed to time-domain data), use the best-fit parameters to simulate an r-FTACV current, and assess the resulting fit to the r-FTACV harmonics. If the result of this process is a single best-fit parameter vector that predicts both sets of data, (i.e. a “good fit to both” in Fig. 2), then these are judged to be the parameters that represent the underlying chemical reality. In other cases (such as a good fit to one form of data but not the other, or when there are multiple competing “best-fit” vectors), the obtained best-fit vector represents a local optimum, and the parameter search needs to be constrained to exclude that region of parameter space. The analysis choices that led to the final best fit, presented below in the section “Resulting parameter inference”, have been formalised into the “recipe” in the next section, for use by the interested experimentalist.

3.1.2. A recipe for parameter inference

The following process is summarised in Fig. 2.

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. Define boundaries for fitting PSV data. Initial boundaries should encompass a reasonably large area of parameter space, but as a rule of thumb should not cover more than two orders of magnitude (and if this scale of coverage is necessary, consider log-transformations). For many parameters, it is possible to obtain an order-of-magnitude estimate from analysing particular current features that are sensitive to the parameter in question. As summarised in Fig. 2, in this paper we obtained order-of-magnitude estimates for $E^0\mu$, (Fig. S5), k_0 (Fig. S10), C_{dl} , (Fig. S7) Γ (Fig. S9) and the phase parameters (Fig. S13). As the exact values of the parameters reported are

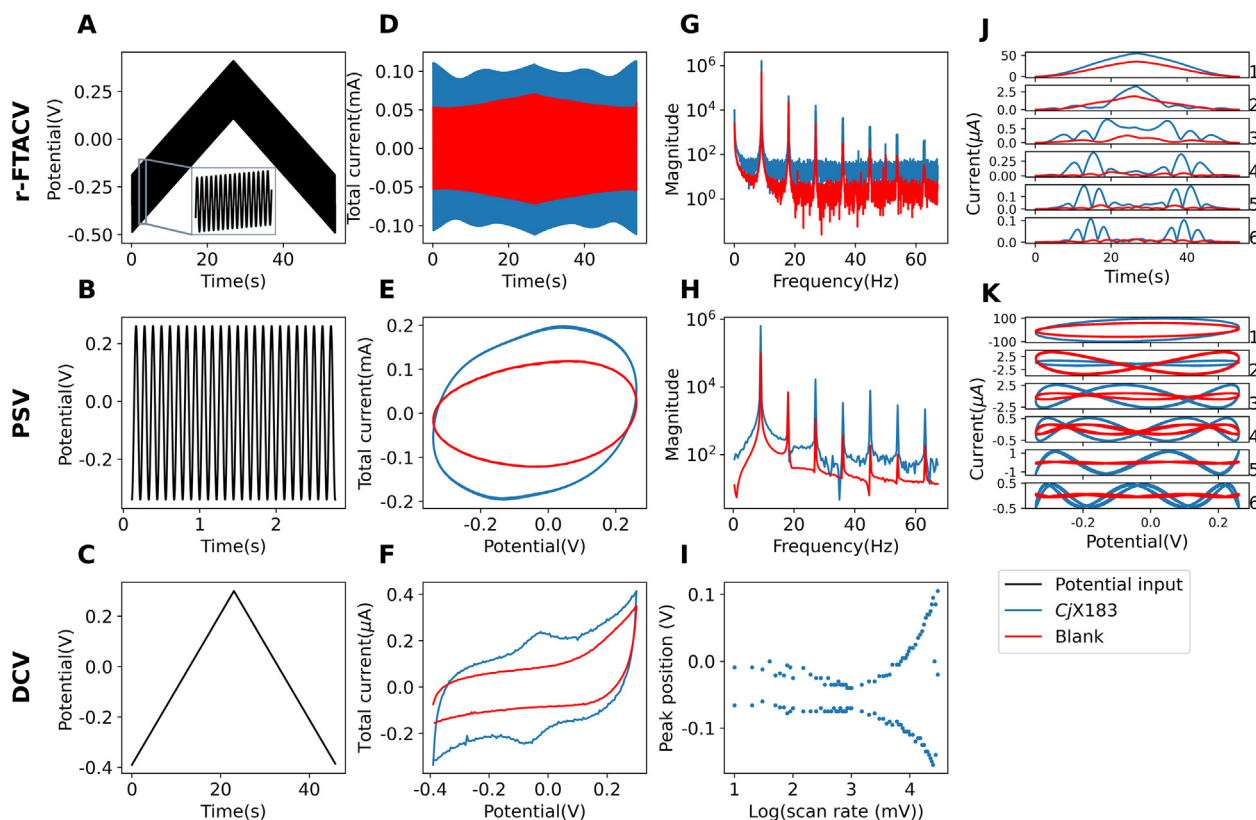


Fig. 1. r-FTACV (top), PSV (centre) and DCV (bottom) experimental data. A-C: Potential inputs for the three experiments. D-F: Total current response of CjX183 (blue) and a bare glassy-carbon electrode (red) to the three experimental inputs. G-H: magnitude of the Fourier spectrum corresponding to harmonics 1–7 for CjX183 (blue) and a bare electrode (red) for r-FTACV and PSV. I: Trumpet plot of DCV peak position data from CjX183 experiments conducted at different scan rates. J-K: r-FTACV and PSV harmonics 1–6 from CjX183 (blue) and a bare electrode (red).

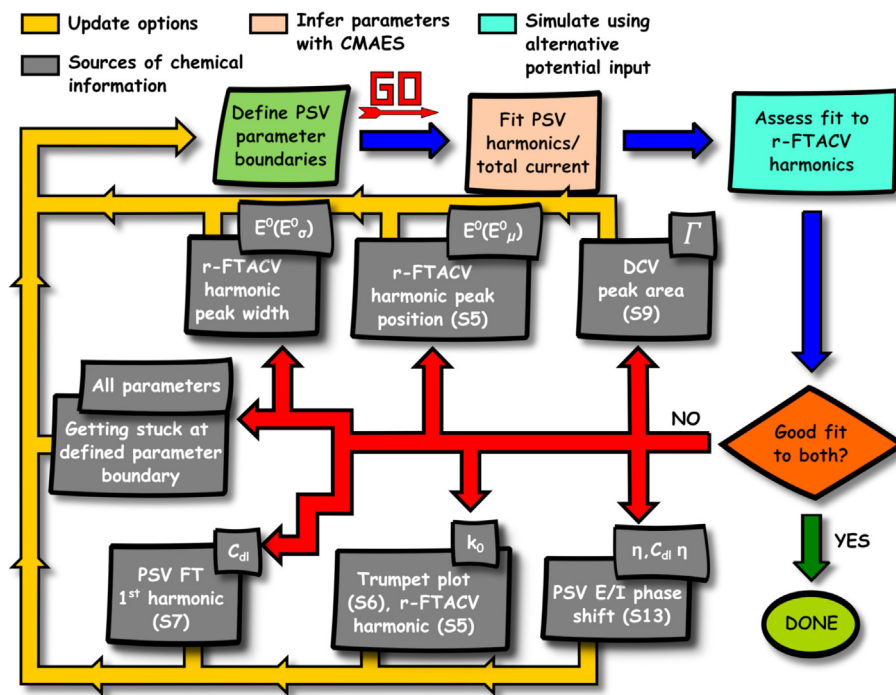


Fig. 2. Flowchart representation of the fitting methodology proposed in this paper. For certain parameters the mechanism by which boundary values can be obtained has been presented graphically; these figures are written in brackets as appropriate.

dependent on the choice of the boundary, this is an extremely important part of the inference process. We describe the reasoning for each bound in the SI.

- 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 (as exemplified in Fig. S3), then it is recommended to fit in the Fourier domain. If this is the case, inspect the harmonics of the blank PSV data (shown in Fig. 1)) to see what portions of the Fourier domain needs to be zeroed-out.
- 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 translatability 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 — this effect is demonstrated in detail in previous work [17]. 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 [17], but this scenario has not been encountered to date.
- Keep on comparing your PSV fits to the r-FTACV harmonics.
 - If filtering of the data is taking place (e.g. excluding the lower harmonics of a PSV experiment), assess how well the predicted current fit the total current — an example of the pitfalls of this approach can be found in Fig. S7.
 - 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 1, the values reported were observed 2–4 times out of ten.
- Choosing which parameter values 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.

As can be gleaned from this recipe and Fig. 2, 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 to return a best-fit parameter set for PSV took approximately 45 min, while the

same attempt with r-FTACV took several hours. As the fitting process is repeated ten times to ensure a high search coverage of parameter space, r-FTACV fitting timescales quickly become untenable. It should be noted that this process is as-described for obtaining point estimates of the parameters, and does not recover parameter distributions. However, once the parameter space is appropriately constrained, these same boundaries can be used to define the prior distribution for Bayesian inference, as discussed below in the Bayesian inference section. This process is fraught with danger, and must be performed as conscientiously as possible.

3.1.3. Constraining parameter space

Constraining parameter space 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, and obtaining false values. The fitting algorithms used in this work require upper and lower bounds as part of their initialisation, and so attempting to pick “reasonable” bounds is the starting point of any fitting process, and these were initially set to prevent only physically implausible values. Because this part of the process can have strong effect on the reported results, the mechanism by which the boundary values are chosen must be informed by the data as much as possible. For several parameters ($E^0\mu$, $E^0\sigma$, Γ , C_{dl} , k_0 and $\eta/C_{dl}\eta$) there exist particular features of the experimental current response that can be used to estimate the values of these parameters, which can in turn be used to define boundaries, as shown in Fig. 2. Additionally for any parameter, as shown in Fig. 2, if the inverse problem solver consistently returns the value of a parameter boundary, this implies that the boundary should be raised or lowered as appropriate, as long as this does not clash with existing chemical knowledge. This process is iterated until a good fit to both PSV and r-FTACV harmonics has been achieved. The final bounds used for each parameter when fitting experiments 1–3 are shown in Table 1. For brevity the reasoning for each parameter bound has been placed in the SI, but the authors recommend that those wishing to use these methods should read these justifications with care before beginning an inference attempt.

3.2. Resulting parameter estimates

Using the bounds shown in Table 1, it was possible to infer parameters from three experiments interrogating different experimental datasets for CjX183, where for each preparation both PSV and r-FTACV measurements were obtained. The resulting best fits are shown in Fig. 3, with the inferred best-fit parameters shown in Table 1, where each column shows data from a different electrode film. In Figs. 3A-C, A-C, PSV harmonics 4–10 for both the experimental and simulated cur-

Table 1

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

Parameter	Symbol	Bounds	Experiment 1	Experiment 2	Experiment 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 (0)	4.70 (0)	4.71 (0)
Phase	Phase (rads)	[3.77, 5.65]	4.57 (0)	4.60 (0)	4.63 (0)
Symmetry factor	α	[0.4, 0.6]	0.6	0.6	0.6

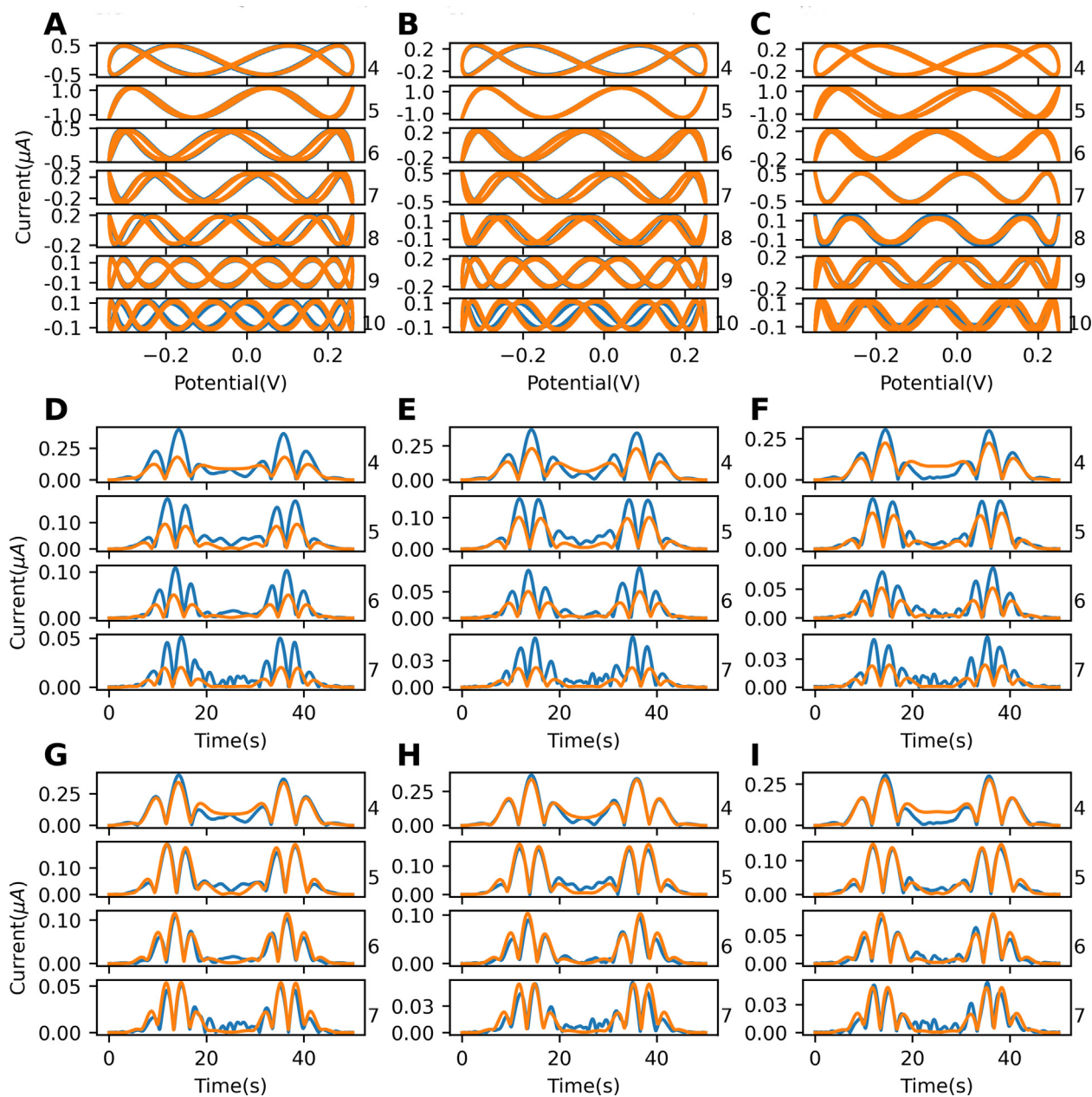


Fig. 3. Best-fit simulations and data for harmonics 4–10 of PSV experiments 1–3 (figures A–C), using simulation parameters reported in Table 1. Figures D–F show r-FTACV harmonics 4–7, from r-FTACV experiments 1–3 (performed using the same CjX183 modified electrode as the appropriate PSV experiment), 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 1.

rents are shown, with the simulations using the parameters in Table 1. In Figs. 3D–F r-FTACV simulations also use the input simulation parameters written in Table 1, except for the input frequency and phase. The sinusoidal frequency ω was fitted, because this is different for r-FTACV and PSV experiments, and the phase was set to the experimentally defined value of 0. For Figs. 3G–I, G–I, limited optimisation was performed on parameters thought likely to change as a result of so-called desorption/inactivation “film-loss” effects, i.e. the Γ parameter, which reflects the number of adsorbed protein molecules, was allowed to vary and so were the so-called “dispersion” E_{μ}^0 and E_{σ}^0 parameters which were optimised based on the assumption that CjX183 molecules adsorbed in certain orientations will be lost from the electrode at a more rapid rate than molecules bound via different surface interac-

tions. The altered values are written in brackets in Table 1. This was done to show that only four of the model input parameters need to be re-optimised in order 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 only requires searching in four-dimensional parameter space (with all other parameters held constant at the values reported in Table 1), 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 limited optimisation fitting to r-FTACV harmonic data, took 2–3 h. It is predicted that fitting r-FTACV data would take 10 + hours when fitting in all dimensions.

The multiple comparison approach for validating the “best-fit” reaction model parameters using both inter-technique and inter-exper-

iment comparisons, such as the type performed in Fig. 3, enables removal of spurious parameter combinations, or the identification of parameter values that only provide a good fit to one subset of the data. Consequently the best-fit parameter values inferred from PSV data that with a small level of alteration provide a good fit to r-FTACV harmonics obtained during the same set of measurements are believed to represent the true underlying redox chemistry of CjX183. This reinforced by the consistency of the values obtained across different experiments performed on different days. Furthermore, our hypothesis that these are “true” protein redox reaction 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 surface coverage estimates from analysis of a single scan-rate DCV experiment. DCV analysis is discussed in more detail in the SI and Fig. S8 including charge calculated from peak area and trumpet plot data [26,9–12], shown in Figs. S9 and S10 respectively. In one sense, it is unsurprising that there is agreement between these PSV and DCV values, as the latter were used to derive the bounds for the former. However, these boundaries were drawn quite broadly (for example, the k_0 boundaries were 50–500 s^{-1} , as informed by a DCV estimate of $\sim 172 \text{ s}^{-1}$), and consequently, there is utility in comparing the PSV and DCV parameter estimates. With regards to the trumpet plot, 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 Fig. S10, simulating trumpet plot data with the PSV-inferred value of $\sim 172 \text{ s}^{-1}$ does not significantly alter the peak-position divergence point. This value is fully consistent with the range of values determined in previous electrochemical studies of different cytochrome proteins. At the lower end, 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} [42]. In the same regime as CjX183, the electron transfer rate constant for human, bovine and porcine cytochrome P450c17 on a PGE electrode were determined to be 164, 157 and 153 s^{-1} respectively [43], and at the upper end, 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} . [44].

With regards to the DCV-estimated surface coverage, although the exact value inferred from the peak-integration technique is dependent on the subtraction approach (in the range of 1.8–3.1e-11 mol cm^{-2}), the values inferred from PSV are in agreement with the range of values inferred by this method. 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 CjX183 on the electrode can be calculated as 4.4 pmol cm^{-2} . Although 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]. The parsimonious experimentalist may conclude that DCV analysis alone can yield similar insights, using both trumpet-plot and single-scan rate analysis. However, it should be noted that DCV-obtained estimates are less reliable. In the case of the trumpet plot, for example, the best-fit simulation did not accurately capture the rate of redox peak divergence with scan rate, and the constant redox peak separation observed at very low scan rates. Additionally, the R_u parameter has to be fixed when fitting trumpet plot data, otherwise the k_0 parameter becomes unidentifiable. The authors believe these issues are primarily caused by challenges associated with background subtraction of capacitance required for DCV analysis, as detailed extensively in the SI for the single scan-rate case. These issues are at least partially resolved by comparison to other techniques, hence the development of a multi-experiment approach. The uncertainty around modelling DCV means

that its primary role is to provide initial order-of-magnitude estimates that inform on how to bound parameter space.

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 timescale of the experiment), the effect of the symmetry factor is low. To explore this effect further, in the SI, in Table S6 and Fig. S14 show the effect of holding the α parameter constant in the range 0.5–0.6 while fitting the other parameters. 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. On the basis of this analysis, the true value of α is likely to reside in the region between 0.55–0.6; the value of the symmetry factor for $\text{Fe}^{2+/3+}$ reactions has consistently been reported to be in the region of 0.4–0.6, and consequently values above 0.6 were considered to be outside of the region of chemical plausibility [45].

In terms of the non-Faradaic parameters, looking at the linear capacitance parameter values in Table 1, the values 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, *vide infra*), 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.

3.3. Bayesian inference

Given that an optimum in parameter space has been found (after parameter space was appropriately constrained), Bayesian inference can be used to recover the probability that a particular set of parameter values around this optimum describe the observed data: the posterior parameter distribution. These are obtained using the adaptive Markov-Chain Monte Carlo algorithm, with 30000 simulations of the forward model generated in a little over an hour. Specific implementation details can be found in other work [46,47]. Three independent chains were run for 10,000 iterations, starting from the best-fit values in Table 1, and the resulting samples used to generate histograms (corresponding to the frequency with which the binned parameter values have been observed) that approximate the parameter posterior distribution for each parameter. Fig. 4, presents these histograms for key parameters obtained from fitting PSV harmonics 4–10 from experiments 1–3 as presented in Fig. 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 extracted from PSV analysis as they both have the same effect on the appearance of the total current [17]). 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, and these were instead fixed at the appropriate value listed in Table 1. It was not practically feasible to undertake a Bayesian inference analysis for r-FTACV, because of the high computational cost of

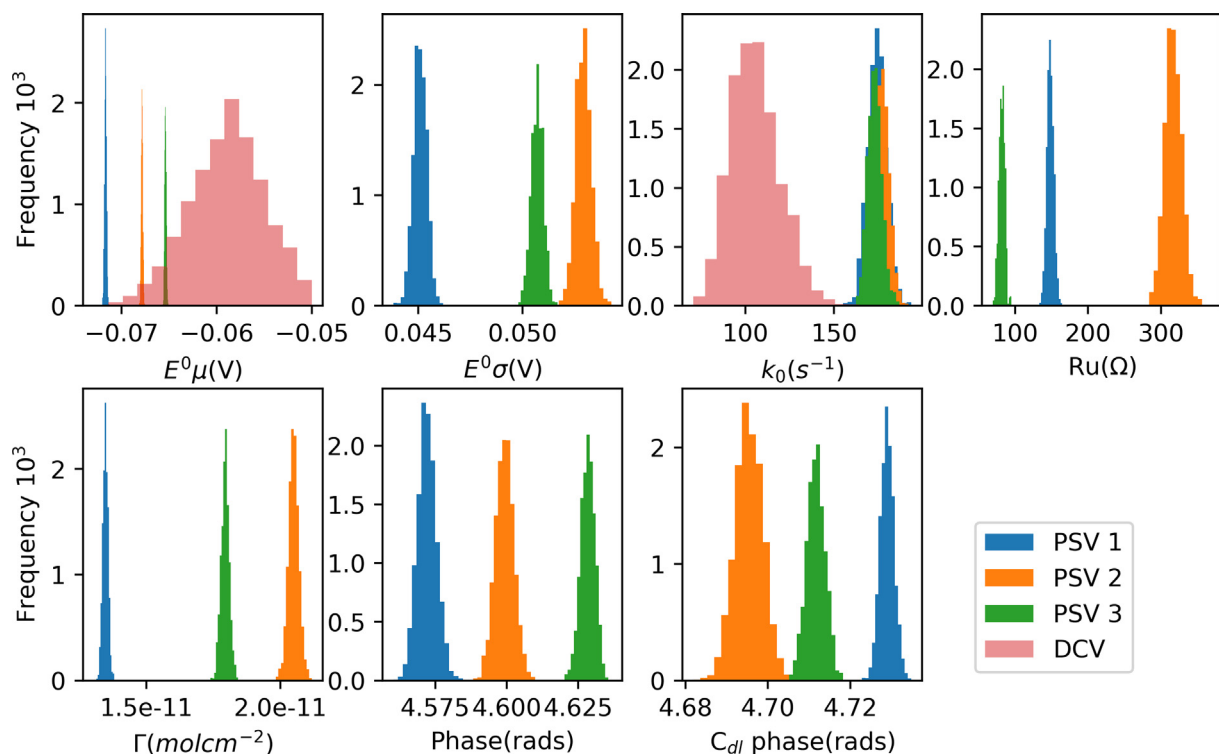


Fig. 4. Inferred parameter histograms generated by pooling three independent MCMC chains, discarding the first 6000 samples as burn-in, using harmonics 4 and above of the respective PSV experiments in the likelihood function, and using the parameters in Table 1 as the starting point. C_{dl} and α were 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 Fig. S6s, and as such only values for the parameters E^0 and k_0 were inferred.

the multiple ~ 10 + hour fitting runs required for a many-parameter MCMC. Fig. 4 demonstrates that despite the above discussed issues of parameter correlation 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. It should be noted that there is a very good agreement between the predicted “underlying” distribution of dispersed E^0 values, which is not immediately apparent from inspection of Fig. 4. These underlying distributions are presented in Fig. S15 in the SI.

In Fig. 5, 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 Figs. S16 and S17. 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 pairwise 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 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 value of 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 the reaction model parameters there are correlations between k_0 , the uncompensated resistance, phase, C_{dl} phase and C_{dlE_2} . Thus, it is clear how challenging it is to define the electron-transfer rate. The positive correlation between E^0 and Γ explains that the discrepancy between the best-fit PSV and r-FTACV values is driven by these parameter correlation effects. As the two-step fits presented in Fig. S11 in the SI show, an increased

C_{dl} value is associated with lowered E^0 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 and Γ parameters. This again shows the utility of the multiple-experiment approach. As r-FTACV is more sensitive to the Faradaic parameters, it allows for the detection of the slight over-estimation of the E^0 parameter, as described above in Fig. 3.

What Fig. 5 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. This is an inescapable problem with attempting to find a global minimum in high-dimensional parameter space that contains multiple local minima. However, the multi-experiment verification approach allows for the mitigation of this issue. It allows for verification of the inferred model parameters, which do not exhibit the same degree or kind of parameter correlation, which is specific to the form of the 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 allow for checking the out-of-sample predictive power of the inferred parameters. This in turn allows for a deep understanding of the relationships between the various parameters, as with the example of C_{dl} , E^0 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.

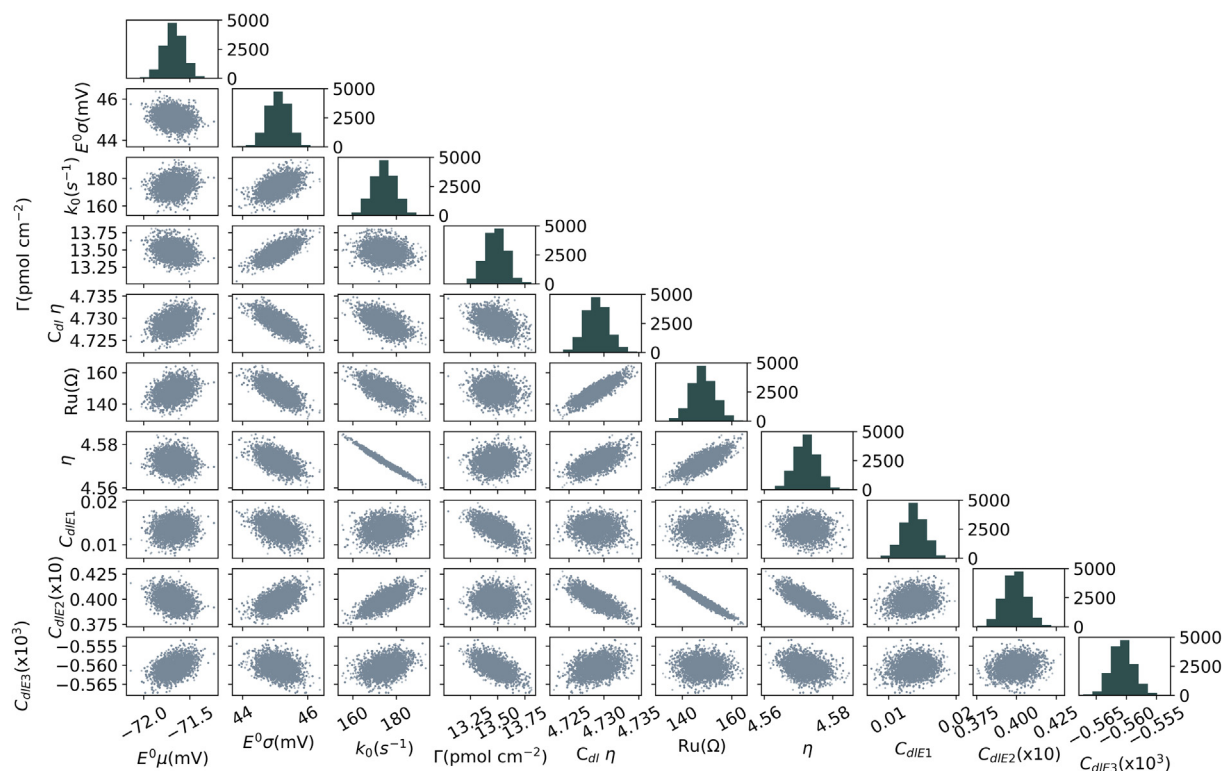


Fig. 5. 2D histograms generated from the MCMC process for PSV experiment 1.

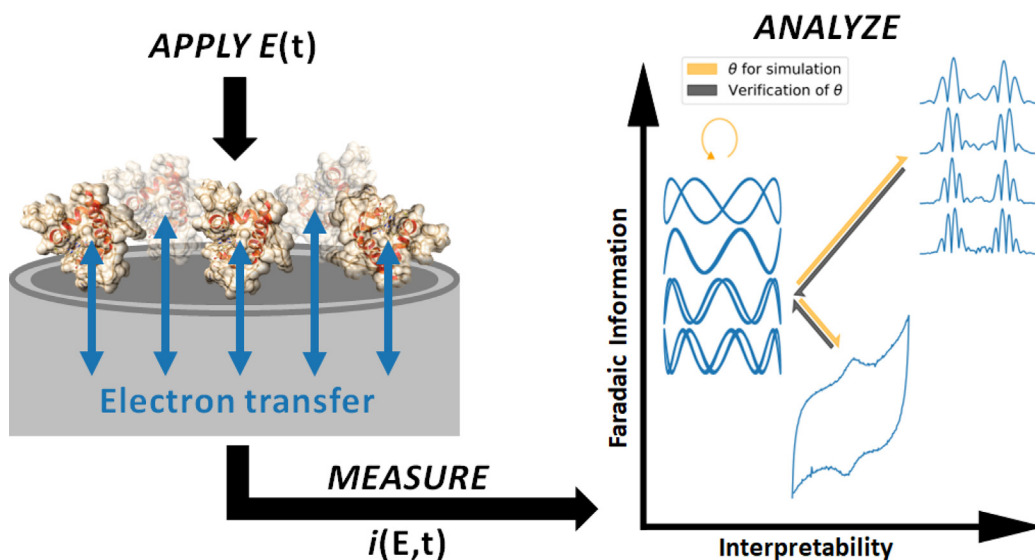


Fig. 6. The three experiments analysed in this paper, direct current voltammetry (DCV) as a current vs. potential plot, purely sinusoidal voltammetry (PSV) current harmonics 4–7 vs. potential, and ramped-Fourier Transform Alternating Current Voltammetry (r-FTACV) current harmonics 4–7 vs. time. The experiments 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.

4. Conclusions

In this work, the inverse problem is solved for PSV harmonics, and the results from the inference process are checked against r-FTACV harmonics, to take advantage of the former's improved simulation speed, and the latter's higher sensitivity to the parameters of interest. The heuristic estimates provided by analysis of DCV data although presenting significant analytical challenges, ultimately prove essential for

appropriate order-of-magnitude bounding of the parameter space searched when fitting PSV data. Furthermore, the relatively rapid simulation time required for PSV ultimately enables the application of Bayesian statistical analysis to the analysis of protein film voltammetry data, allowing for unprecedented insight into the fitting process. A summary of how the techniques relate to one another is provided in Fig. 6. Overall, in terms of the accuracy of the point estimates, parameter sensitivity studies demonstrate that r-FTACV is the most sensitive

of the three techniques [48], and consequently the parameter vector which provides a good fit to the r-FTACV harmonics is judged to be the best representation of the redox chemistry of CjX183. There is a small discrepancy between these parameters and the best-fit PSV parameters; the source of this discrepancy is parameter compensation, as uncovered by MCMC, and as such these values are judged to be slightly less accurate. Finally, DCV is primarily used for order-of-magnitude estimates for key parameters, because of the challenges of background subtraction.

The overall goal of this work is to obtain an understanding of protein bio-electrochemistry by extracting inferred redox reaction parameters. To achieve this, the parameter inference process needs to be rapid, accurate and reproducible. We have described how this can be facilitated by complementing protein-film PSV analysis, which makes it possible to obtain parameter estimates on a short timescale, with parameter validation based on DCV and r-FTACV analysis. Bayesian inference, also facilitated by the rapid simulation speed of PSV, allows for an understanding of how parameters compensate for each other during the fitting process. It is hoped that such approaches will become an indispensable component of analysing protein film voltammetry data in the future, and we have listed a detailed “recipe” for our methodology in the SI.

In 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 chemistry, including multiple electron-transfer reactions and catalytic processes.

CRedit authorship contribution statement

Henry O. Lloyd-Laney: Conceptualization, Methodology, Software, Validation, Writing - original draft, Writing - review & editing. **Nicholas D.J. Yates:** Investigation. **Martin J. Robinson:** Methodology, Writing - review & editing. **Alice R. Hewson:** Investigation. **Jessie Branch:** Resources. **Glyn R. Hemsworth:** Resources, Writing - review & editing. **Alan M. Bond:** Conceptualization, Writing - review & editing. **Alison Parkin:** Conceptualization, Methodology, Writing - review & editing. **David J. Gavaghan:** Conceptualization, Methodology, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.jelechem.2023.117264>.

References

- [1] H. Adamson, A.M. Bond, A. Parkin, *Chem. Commun.* 53 (2017) 9519–9533.
- [2] O. Lampret, J. Duan, E. Hofmann, M. Winkler, F.A. Armstrong, T. Happe, *PNAS* 117 (2020) 20520–20529.
- [3] R. Gulaboski, V. Mirčeski, I. Bogeski, M.J. Hoth, *Solid State, Electrochem.* 16 (2012) 2315–2328.
- [4] C. Léger, P. Bertrand, *Chem. Rev.* 108 (2008) 2379–2438.
- [5] D.J. Gavaghan, J. Cooper, A.C. Daly, C. Gill, K. Gillow, M. Robinson, A.N. Simonov, J. Zhang, A.M. Bond, *ChemElectroChem* 5 (2018) 917–935.

- [6] V. Fourmond, E.S. Wiedner, W.J. Shaw, C.J. Léger, *Am. Chem. Soc.* 141 (2019) 11269–11285.
- [7] V. Fourmond, C. Léger, *Curr Opin, Electrochem.* 1 (2017) 110–120.
- [8] L. Bieniasz, B.J. Speiser, *Electroanal. Chem.* 458 (1998) 209–229.
- [9] R.C. Stan, A. Kros, N. Akkilic, N.J. Sanghamitra, J.J. Appel, *Electroanal. Chem.* 787 (2017) 14–18.
- [10] B.D. Fleming, J. Zhang, D. Elton, A.M. Bond, *Anal. Chem.* 79 (2007) 6515–6526.
- [11] F.A. Armstrong, H.A. Heering, J. Hirst, *Chem. Soc. Rev.* 26 (1997) 169–179.
- [12] F.A.J. Armstrong, *Chem. Soc. Dalton Trans.* (2002) 661–671.
- [13] B. Munge, S.K. Das, R. Ilagan, Z. Pendon, J. Yang, H.A. Frank, J.F.J. Rusling, *Am. Chem. Soc.* 125 (2003) 12457–12463.
- [14] R. Gao, M.A. Edwards, J.M. Harris, H.S. White, *Curr. Opin. Electrochem.* 22 (2020) 170–177.
- [15] M. Lu, R.G. Compton, *Analyst* 139 (2014) 4599–4605.
- [16] C. Léger, A.K. Jones, S.P. Albracht, F.A.J. Armstrong, *Phys. Chem. B* 106 (2002) 13058–13063.
- [17] H.O. Lloyd-Laney, M.J. Robinson, A.M. Bond, A. Parkin, D.J.J. Gavaghan, *Electroanal. Chem.* (2021) 115204.
- [18] H.O. Lloyd-Laney, N.D. Yates, M.J. Robinson, A.R. Hewson, J.D. Firth, D.M. Elton, J. Zhang, A.M. Bond, A. Parkin, D.J. Gavaghan, *Anal. Chem.* 93 (2021) 2062–2071.
- [19] J. Hirst, *BBA - Bioenergetics* 1757 (2006) 225–239.
- [20] F.A. Armstrong, R.M. Evans, S.V. Hexter, B.J. Murphy, M.M. Roessler, P. Wulff, *Acc. Chem. Res.* 49 (2016) 884–892.
- [21] R. Gulaboski, V.J. Mirčeski, *Solid State, Electrochem.* 24 (2020) 2723–2732.
- [22] M. Merrouch, J. Hadj-Said, C. Léger, S. Dementin, V. Fourmond, *Electrochim. Acta* 245 (2017) 1059–1064.
- [23] T. Reda, J.J. Hirst, *Phys. Chem. B* 110 (2006) 1394–1404.
- [24] A.R. Dale-Evans, M.J. Robinson, H.O. Lloyd-Laney, D.J. Gavaghan, A.M. Bond, A. Parkin, *Front. Chem.* 9 (2021) 424.
- [25] A.M. Bond, N.W. Duffy, S.-X. Guo, J. Zhang, D. Elton, *Anal. Chem.* 77 (2005) 186–A.
- [26] E.J. Laviron, *Electroanal. Chem. Interfacial, Electrochem.* 101 (1979) 19–28.
- [27] Mirčeski, V.; Komorsky-Lovrić, Š.; Lovrić, M. In *Square-wave voltammetry: theory and application*; Scholz, F., Ed.; Springer Science & Business Media, 2007..
- [28] A. Lasia, *Electrochemical impedance spectroscopy and its applications*, Springer, 2020.
- [29] L.J. Jeuken, J.P. McEvoy, F.A.J. Armstrong, *Phys. Chem. B* 106 (2002) 2304–2313.
- [30] C.E. Lubner, D.P. Jennings, D.W. Mulder, G.J. Schut, O.A. Zadovornyy, J.P. Hoben, N. Tokmina-Lukaszewska, L. Berry, D.M. Nguyen, G.L. Lipscomb, et al, *Nature Chem. Biol.* 13 (2017) 655–659.
- [31] K. Pandey, S.T. Islam, T. Happe, F.A. Armstrong, *PNAS* 114 (2017) 3843–3848.
- [32] Armstrong, F.A.; Evans, R.M.; Megarity, C.F. *Methods in enzymology*; Elsevier, 2018; Vol. 599; Chapter 14, pp 387–407..
- [33] H. Adamson, M. Robinson, P.S. Bond, B. Soboh, K. Gillow, A.N. Simonov, D.M. Elton, A.M. Bond, R.G. Sawers, D.J. Gavaghan, et al, *Anal. Chem.* 89 (2017) 1565–1573.
- [34] J. Huang, M. Papac, R. O’Hayre, *Electrochim. Acta* 367 (2021) 137493.
- [35] L. Gundry, G. Kennedy, J. Keith, M. Robinson, D. Gavaghan, A.M. Bond, J. Zhang, *ChemElectroChem* 8 (2021) 2238–2258.
- [36] N. Armstrong, D. Hibbert, *Chemom. Intell. Lab. Syst.* 97 (2009) 194–210.
- [37] J. Branch, B.S. Rajagopal, A. Paradisi, N. Yates, P.J. Lindley, J. Smith, K. Hollingsworth, W.B. Turnbull, B. Henrissat, A. Parkin, et al, *Biochem. J.* 478 (2021) 2927–2944.
- [38] L.L. Leggio, T.J. Simmons, J.-C.N. Poulsen, K.E. Frandsen, G.R. Hemsworth, M.A. Stringer, P. Von Freiesleben, M. Tovborg, K.S. Johansen, L. De Maria, et al, *Nat. Commun.* 6 (2015) 1–9.
- [39] M. Clerx, M. Robinson, B. Lambert, C.L. Lei, S. Ghosh, G.R. Mirams, D.J. Gavaghan, *Journal of Open Research Software* (2019).
- [40] H. Adamson, M. Robinson, J.J. Wright, L.A. Flanagan, J. Walton, D. Elton, D.J. Gavaghan, A.M. Bond, M.M. Roessler, A.J. Parkin, *Am. Chem. Soc.* 139 (2017) 10677–10686.
- [41] G.P. Morris, R.E. Baker, K. Gillow, J.J. Davis, D.J. Gavaghan, A.M. Bond, *Langmuir* 31 (2015) 4996–5004.
- [42] S. Jian, X. Liu, H. Sun, S. Hou, *RSC Adv.* 4 (2014) 6165–6172.
- [43] D.L. Johnson, A.J. Conley, L.L.J. Martin, *Mol. Endocrinol.* 36 (2006) 349–359.
- [44] J. Wei, H. Liu, K. Niki, E. Margoliash, D.J. Waldeck, *Phys. Chem. B* 108 (2004) 16912–16917.
- [45] R. Tamamushi, *Kinetic parameters of electrode reactions of metallic compounds*, Butterworths London (1975).
- [46] R.H. Johnstone, E.T. Chang, R. Bardenet, T.P. De Boer, D.J. Gavaghan, P. Pathmanathan, R.H. Clayton, G.R.J. Mirams, *Mol. Cell. Cardiol.* 96 (2016) 49–62.
- [47] H. Haario, E. Saksman, J. Tamminen, Bernoulli (2001) 223–242.
- [48] Lloyd-Laney, H. *Computationally interrogating enzyme electrochemistry for rational mutation of metalloproteins*. Ph.D. thesis, University of Oxford, 2021.