Automated algorithm for measurement of spontaneous adenosine transients in large electrochemical data sets

Ryan P. Borman
De Pere, Wisconsin

Bachelor of Science in Chemistry, University of Colorado-Denver, 2013

A Thesis presented to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of
Master of Science

Department of Chemistry

University of Virginia
July, 2016

_____

_____

_____

_____

Automated algorithm for measurement of spontaneous adenosine transients in large electrochemical data sets

Abstract

Spontaneous adenosine release events have been discovered in the brain that last only a few seconds. The identification of these adenosine events from fast-scan cyclic voltammetry data has been performed manually and is difficult due to the random nature of adenosine release. In this study, we develop an algorithm that automatically identifies and characterizes adenosine transient features, including event time, concentration, and duration. Automating the data analysis reduces analysis time from 10-18 hours to about 40 minutes per experiment. The algorithm identifies adenosine based on its two oxidation peaks, the time delay between them, and their peak ratios. In order to validate the program, 4 data sets from 3 independent researchers were analyzed by the algorithm and then verified by an analyst. The algorithm resulted in 10 ± 4% false negatives and 9 ± 3% false positives. The specificity of the algorithm was verified by comparing calibration data for adenosine triphosphate, histamine, hydrogen peroxide, and pH changes and these analytes were not identified as adenosine. Stimulated histamine release in vivo was also not identified as adenosine. The code is modular in design and could be easily adjusted to detect features of spontaneous dopamine or other neurochemical transients in FSCV data.

Table of Contents

List of Illustrations:

Acknowledgement

I would like to acknowledge my advisor, Dr. Jill Venton, for helping me grow as a scientist by thinking more critically about my work and by teaching me how to communicate my ideas better. I also would like to acknowledge everyone one in my lab, whom I bothered with endless amounts of questions and always took their time helping me understand the answers. I would especially like to thank Dr. Michael Nguyen for entertaining this hair-brained idea of automating adenosine detection. It started as a joke (and now it's a thesis) because we were not experts in computer programming. Also, I would like to thank Dr. Ashley Ross for editing many of my papers and helping me develop a clever experiment for this thesis. Lastly, I would like to think Ying Wang and Scott Lee for taking their time verifying the validity of the program with their data.

Ultimately, I would like to thank my parents, John and Julie Borman, who have been very supportive of my decisions and I could not ask for better parents. When I was down, they picked me up and always kept things in perspective.

Chapter 1: Introduction

1.1. Adenosine overview

1.1.1 Adenosine biological role

Adenosine is an important biological nucleoside that is involved in many metabolic processes including cell signaling[1], neuromodulation[1,2], and neuroprotection[1,2]. Adenosine is a constituent of all cells since it is a byproduct of ATP catabolism[1] and is found in many regions of the brain including the caudate-putamen[3], hippocampus[4], nucleus accumbens[3], and cortex[5]. Endogenous adenosine acts as a neuromodulator and plays an active role in the regulation of cerebral blood flow[6,7]. Neuromodulators have been typically thought to act on slow time scales, minute to hours, by volume transmission. Volume transmission happens within the brain extracellular fluid and can include both short and long distances[8]. Adenosine modulates by binding to one of four G protein-coupled receptors to cause either an inhibitory or stimulatory regulation of neurotransmitters [6]. Release of adenosine has been shown to protect heart cells during ischemia, where there is a deficiency in oxygen delivery[9]. The ability to measure adenosine is important in understanding the role it plays in neuromodulation and homeostatic regulation. Furthermore, since it is pervasive throughout the central nervous system and found in every cell, understanding the function adenosine plays in physiological disorders like hypoxia and ischemia could help with new treatments for these disorders. Adenosine traditionally has been studied as a slow acting molecule but recent evidence suggests a more rapid spontaneous mode of signaling[7]. The discovery of rapid spontaneously released adenosine in the brain on the minute timescale has generated a need for automatic feature detection due to hundreds of events being released in a single animal experiment. In this study a straightforward algorithm was

designed to identify and characterize random adenosine transients from FSCV color plots.

1.1.2 Adenosine production

The production and regulation of adenosine is a highly complex system involving both intracellular and extracellular formation mechanisms (*Figure 1.1*). Adenosine can be formed intracellularly in the central nervous system by the catabolism of AMP, due to metabolic stress, or by cytosolic 5'-nucleotidase. The intracellular formation of
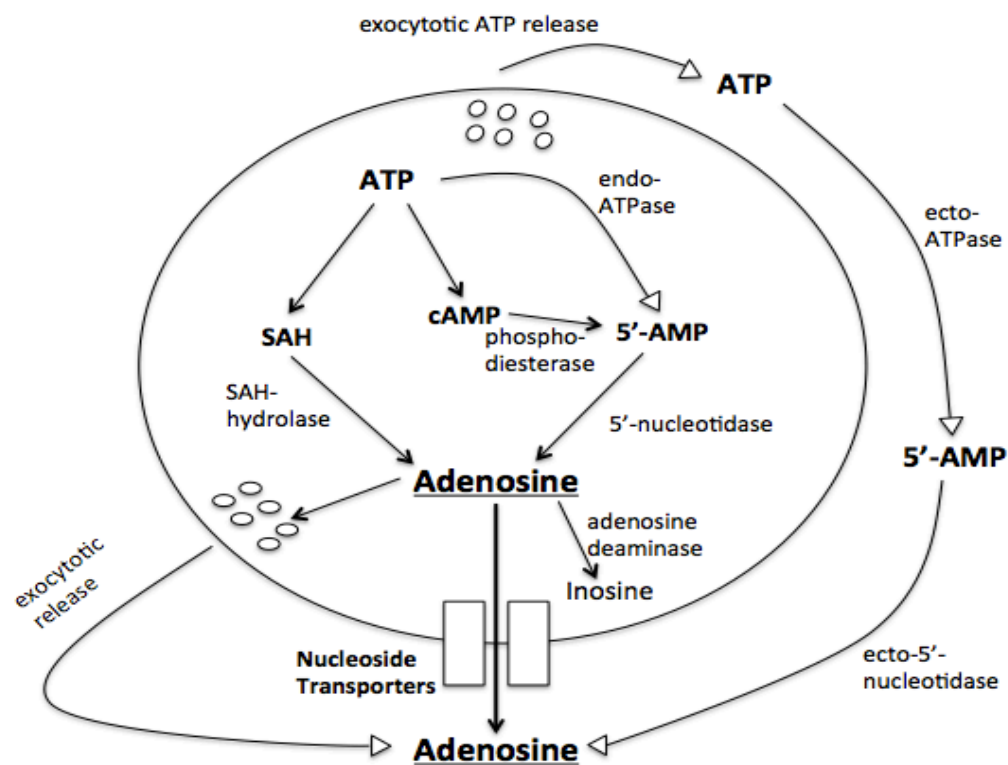

Figure 1.1: Mechanism for adenosine production. Intracellularly adenosine can be formed from ATP. Alternatively, adenosine can be formed from metabolism of ATP extracellularly. Modified by Pajeski *et al.* [3]

adenosine can also be formed from the hydrolysis of s-adenosylhomocysteine (SAH in *Figure 1.1*)[2]. Adenosine can be released to the extracellular space by two different

mechanisms: bi-directional nucleoside transporters or exocytosis of ATP that metabolizes to adenosine.

1.1.3 Adenosine regulation and function

There are four adenosine receptors expressed in the brain, which are A_1, A_{2A}, A_{2B}, and A_3. A_1 and A_{2A} are high affinity for adenosine with binding affinities in the 1-30 nM range[10]. A_1 and A_{2A} receptors are thought to be active at normal physiological conditions since the basal level of adenosine is at low nanomolar concentrations[1]. The A_1 receptor is an inhibitory G-protein coupled receptor (G_i). Alternatively, A_{2A} is an excitatory G-protein coupled receptor (G_e). The A_{2B} and A_3 receptors have lower binding affinities; in the 1-20 µM range[11]. The activation of A_{2B} and A_3 receptors is thought to occur under stressful physiological conditions like hypoxia or ischemia due to higher concentrations of adenosine[6]. Overall, adenosine is a highly complex signaling molecule in the brain and many questions are still unanswered on how adenosine is regulated.

1.1.4 Adenosines physiological role

Adenosine modulates numerous important physiological functions including sleep[12], breathing[13], and heart rate[14]. Furthermore, adenosine is directly involved in pathologies like inflammation and cerebral ischemia. A_{2A} adenosine receptors increase immunosuppressive cAMP in the immune cells of mice and play a role in the attenuation of inflammation and tissue damage *in vivo*[15]. Moreover, adenosine has been studied for it role as a neuroprotectant against damage in cerebral ischemia and as a possible therapeutic for stroke[16]. In response to energy depletion induced by ischemia, extracellular concentrations of adenosine can increase 1000 fold. During pathological events like cerebral ischemia and inflammation the increase of adenosine typically lasts

minutes to hours[5]. Recently, adenosine was shown to rapidly modulate stimulated dopamine release in the caudate-putamen by A_1 receptors[8].

1.1.5 Adenosine release

Metabolic processes and the breakdown of ATP during energy consumption can cause a buildup of adenosine in the extracellular space. Typically, these extracellular adenosine concentrations have been studied use techniques with only minute to hour temporal resolution. One group used electrophysiological techniques to explore a rapid modulatory role for adenosine in the brain[4]. Electrically stimulated adenosine regulates glutamate receptor-mediated excitatory postsynaptic potentials (EPSPs) in the hippocampus. The measured duration of the EPSPs were 2 seconds, suggesting that adenosine causes changes at sub-minute timescales. This suggests that traditional measurement of adenosine, formerly delegated as a slow acting molecule, does not sufficiently describe adenosines role in rapid signaling.

Subsecond adenosine changes have also been directly measured from electrical stimulations in striatal rat brain slices by fast-scan cyclic voltammetry[17]. The results suggest that adenosine release is activity-dependent. Stimulated adenosine has been studied *in vivo* using carbon-fiber microelectrodes as adenosine sensors[6]. The purpose of this experiment was to determine stimulated adenosine release in rat caudate-putamen, after electrical stimulation. Results show that adenosine increased in the extracellular space and was cleared in about 15 seconds.

Recently, a new form of adenosine release has been found for the first time[5]. An adenosine transient is a non-stimulated occurrence of adenosine with event duration on the order of seconds. Spontaneous release of adenosine was measured in rat caudate-putamen and the prefrontal cortex. Average concentrations of adenosine release were

0.18 µM and had a range of 0.04-3.2 µM, in both brain regions. The study illustrates that adenosine is rapidly released and cleared in the brain, which suggests that adenosine is involved in rapid neuromodulation in addition to the longer term neuromodulation described above. The frequency of spontaneous adenosine events for a single transient was every 2–3 minutes. This study demonstrates that adenosine events are spontaneous and do not follow any regular pattern and therefore are random events[5].

Due to these adenosine transients being spontaneous and random, an analyst must find each adenosine transient by "hand", which is very time consuming. Depending on the type of experiment (i.e. brain slice/*in vivo* models and pharmacological/stroke experiments), 2–4 hours of data are obtained and the number of adenosine transients varies widely. Furthermore, the more transients in a data set the longer time it takes for data processing. A user who is experienced in data analysis can find one adenosine transient approximately every 1.5 minutes, therefore a data set containing 700 transients will take a user 18 hours to analyze. Automation of the data analysis process will allow an analyst to more than double their experimental production when high transient counts are measured.

1.2. Techniques for measuring adenosine

1.2.1 Measuring adenosine by microdialysis

Historically, microdialysis coupled with HPLC has been one of the most employed techniques for measuring adenosine[12,18,19]. Microdialysis is a sampling technique that is used frequently in neurobiology because it is minimally invasive, has the ability to sample continuously, and measures basal levels of analytes. During ischemia, acutely implanted microdialysis probes (300 µm in outer diameter), measured an increase in the dialysate levels of adenosine compared to chronically implanted

probes[19]. Acute and chronic measurements of adenosine were taken at 2 and 24 hours, respectively. Microdialysis probes are relatively large and therefore can disrupt the accuracy of the measurement of adenosine.

1.2.2 Measuring adenosine with biosensors

Enzyme based sensors are used for measuring adenosine at sub-minute temporal resolution. The Dale group developed a three enzyme biosensor for the detection of adenosine[20]. It works by breaking down adenosine to inosine via adenosine deaminase, subsequently to hypoxanthine via purine nucleoside phosphorylase, finally to xanthine, urate, and hydrogen peroxide via xanthine oxidase. These enzymes work to metabolize adenosine to a final product of hydrogen peroxide, which is amperometrically detected. Adenosine biosensors are 25 – 100 μm in diameter, have a detection limit of 12 nM, and have a temporal resolution of 2 seconds. A null sensor that does not contain adenosine deaminase is positioned next to the biosensor to measure any interfering downstream metabolites. The null detector signal can be subtracted from the biosensor to obtain the biosensors response to adenosine[7]. Adenosine biosensors are smaller than typical microdialysis probes, have low detection limits, and sub-minute temporal resolution. An alternative to biosensors for the detection of adenosine is fast-scan cyclic voltammetry, an electrochemical technique that directly measures adenosine at carbon-fiber microelectrodes.

1.2.3 Measuring adenosine by fast-scan cyclic voltammetry

Fast-scan cyclic voltammetry (FSCV) is an electrochemical technique that was developed to measure real-time changes in dopamine levels *in vivo*[21]. FSCV is similar to traditional cyclic voltammetry but FSCV has faster temporal resolution. With FSCV, a carbon-fiber microelectrode (CFME) is scanned from a negative holding potential to a positive switching potential and immediately ramped back down. (*Figure 1.2*) The total length of the scan is about 10 ms, and data is collected at 10 Hz, which gives 100 ms temporal resolution. Background charging currents are stable with CFMEs, which allows for accurate background subtraction. Charging currents result from the formation of a double layer of ions at the CFME interface that act similar to a capacitor[22]. Background subtracted cyclic voltammograms provide a selective fingerprint for each analyte of interest. The CFME has a diameter of 7 μm, which allows it to be placed in specific brain regions while minimizing potential tissue damage compared to larger microdialysis electrodes or enzyme biosensors. An advantage of FSCV is its ability to measure rapid changes in electroactive neurotransmitters. To measure adenosine by FSCV a triangle waveform is applied to a CFME that scans from a holding potential of -0.4 V to a switching potential of 1.45 V versus a Ag/AgCl reference electrode at a scan rate of 400 V/s. The limit of detection for fast-scan cyclic voltammetry at a carbon-fiber microelectrode is 15 nM, which is similar to detection limits at enzyme biosensors of 12 nM[7].
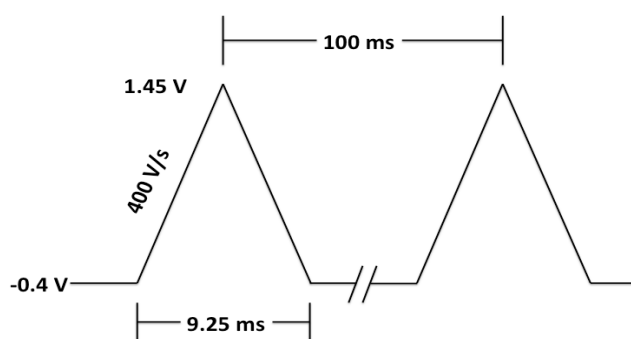

Figure 1.2: The applied potential waveform for *in vivo* adenosine measurement. The waveform is scanned from -0.4 V to 1.45 V at a scan rate of 400 V/s. Each scan is 10 ms and scans are repeated every 100 ms

Adenosine is an electroactive molecule that can go through three successive, two-electron oxidations (*Scheme 1.1*)[23]. When a triangle waveform is applied to a CFME, an adenosine molecule undergoes a two-electron primary oxidation to form product II in *Scheme 1.1*. This primary oxidation product is observed at 1.4 V. Subsequently, a secondary oxidation occurs to form product III at 1.0 V. These first two oxidation steps are irreversible and no reduction peak is observed in its corresponding cyclic voltammograms (CV) as can be seen in *Figure 1.3B*. Typically, the tertiary oxidation product III is not detected using FSCV at our CFMEs.
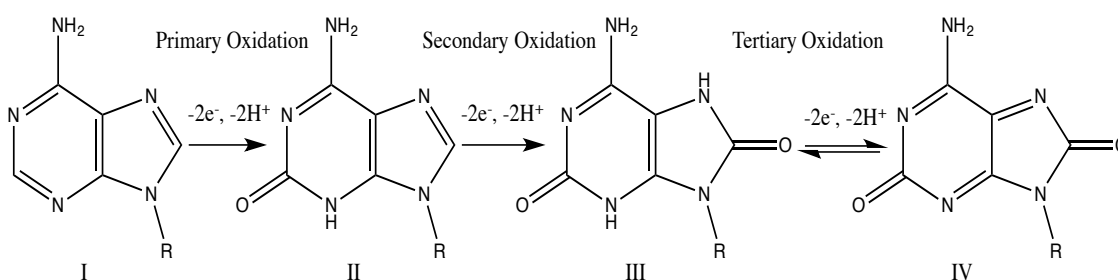
Scheme 1.1: Adenosine (I) undergoes a two-electron primary oxidation at 1.4 V to form product II. Subsequently, product II is involved in a secondary oxidation at 1.0 V to form product III. Product IV is normally not observed in FSCV. In this scheme R is ribose.

The oxidation of adenosine at our carbon-fiber microelectrodes is a two-step process. The primary product is produced in an irreversible oxidation and further oxidizes to a secondary product at our CFMEs[5]. This can be visualized in a color plot of *in vivo* spontaneous transient adenosine release (*Figure 1.3A*). During FSCV cyclic voltammograms are taken 10 times per second. A color plot displays cyclic voltammograms as a function of time with current represented in false color. As can be visualized the color plot, the start of the primary peak and secondary peak are not at the same time point. The dashed lines in the color plot represent times where CVs were taken for adenosine at three different time points (*Figure 1.3B*). The first CV, taken at the start of the primary peak, shows the primary peak forming and the secondary peak absent. In the second CV, taken at the primary peak maximum, the secondary peak has

started to form. In the third CV, the primary peak is waning and the secondary peak is at a maximum. Adenosines signal changes as a function of time with the secondary peak evolving only after the primary peak is produced. The lag time between the primary and secondary peak max can be utilized to identify spontaneous adenosine transients.
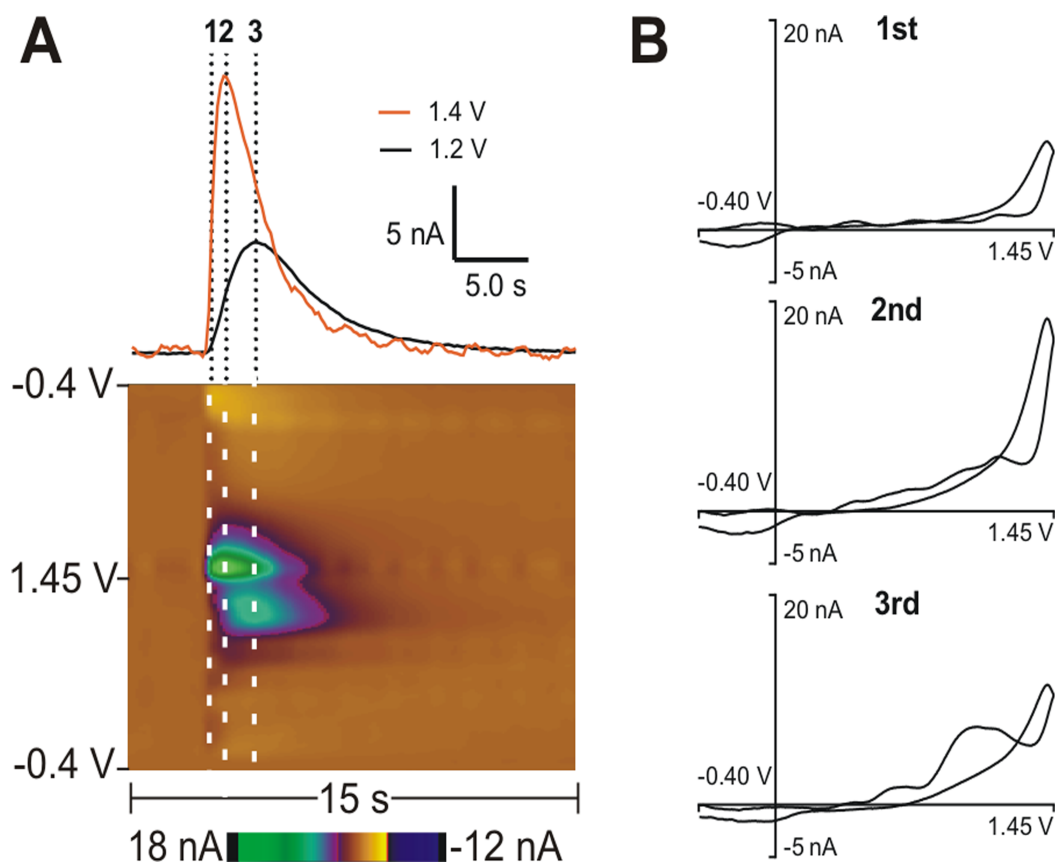
Figure 1.3: *in vivo* Spontaneous transient release of adenosine. (A) The top shows current versus time traces of adenosine at 1.4 V for the primary oxidation (orange) and 1.2 V for the secondary oxidation (black). The bottom is a false color plot with dashed lines showing where the CVs were taken. (B) Cyclic voltammograms taken at different time intervals. The top CV was taken at the start of the primary oxidation, the middle CV was taken at the primary peak maximum, and the bottom CV was taken at the secondary peak maximum. Figure taken from Nguyen *et al.* [5]
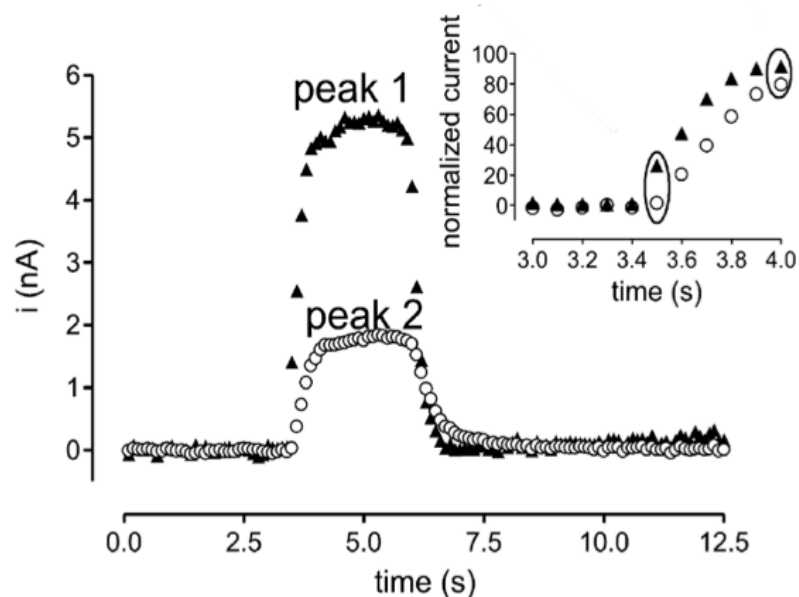
Figure 1.4: Current versus time traces for primary peak, taken at 1.5 V, in triangles, and secondary peak, taken at 1.0 V, in circles of adenosine in flow injection analysis experiment. The rise of the primary peak occurs before the rise of secondary peak. *Inset* : At 3.5 seconds the rise of the primary peak starts and the rise of the secondary peak starts at 3.6 seconds. Therefore, the primary oxidation product must be produced before the secondary product is able to be formed. Figure taken from Swamy *et al.* [9]

Adenosines primary oxidation peak starts before the secondary oxidation peak is produced (*Figure 1.4*). This is a necessary condition for the identification of adenosine at CFMEs and basis is established in the redox chemistry of adenosine[24]. Current versus time traces of the primary (Peak 1, triangles) and secondary (Peak 2, circles) peak maximums, in *Figure 1.4*, show a lag time between the production of the first and second oxidation product. In the inset of *Figure 1.4* a plot of normalized current versus time shows a rise in primary peak at 3.5 seconds and the secondary peak rising at 3.6 seconds. In large data sets it is very time consuming to identify adenosine transients and many hours are devoted to identifying and characterizing these transients in our lab. The oxidation voltages for the primary and secondary peaks are specific to adenosine and the difference in time between these peaks is a way to identify adenosine.

1.3. Automating adenosine identification

The ability to automate recognition of spontaneous adenosine transients would have two major advantages: (1) avoiding large amounts of time and labor involved in data analysis and (2) making data analysis consistent within and between different experiments and researchers. It takes about 1.5 minutes for an experienced analyst to identify an adenosine transient and a typical animal experiment can have as many as 700 transients. It is possible that multiple researchers, performing data analysis on a single data set, could arrive at different conclusions with that data. It is more likely that data analysis is performed on different experimental data and researchers are interested in comparing independent data sets. Cyclic voltammograms are ideal for automating identification of adenosine transients since adenosine's oxidation potentials for primary and secondary peaks are at specific voltages. Having a fully automatic computer aided algorithm that determines adenosine features like event time, concentration, and duration of adenosine transients, will enable researchers to compare differing data sets and draw more accurate conclusions upon these data with less time spent counting transients by hand.   One existing technique for peak identification is principal components regression (PCR) but it is not fully automated in identifying transient peaks and their features.

1.3.1 Principal components analysis

Principal components analysis (PCA) is a multivariate statistical technique that reduces dimensionality by retaining relevant and discarding non-relevant information provided in large data sets. The combination of principal components analysis with inverse least-squares regression is known as principal components regression (PCR)[25]. PCA is used to determine relevant information from non-relevant noise. PCR then can

use residual analysis and remove noise from unknown data. Succinctly, residuals are the difference between an observed value and an estimated value, which is the value of interest. If the summed square of residual current at any applied potential of a CV (Q_t) exceeds the threshold value in a training set (Q_α) then a source of variation is not accounted for in the PCA model and the value should be retained[25,26]. An example of PCR can be observed in *Figure 1.5* for dopamine and pH. When dopamine and pH are included in the PCR training set they are removed from the residual currents as seen in *Figure 1.5B*[27]. Essentially the CVs included in the training set are removed, which suggests the training set accurately describes dopamine and pH. This is a way of showing how well a training set's PCs describe the unknown data. The Q trace in *Figure 1.5C* displays no significant current contributions other than dopamine and pH in the residual plot. The dotted line in *Figure 1.5C* refers to the threshold Q_α at the 95% significance level. The resulting color plot in *Figure 1.5B* can be attributed to noise and be removed from the original color plot. Similarly, if only dopamine is in the training set, the residual color plot and Q trace displays pH only, which can be seen in *Figure 1.5D* and 1.*5E*, respectively. PCR is a good method for removing spurious noise and interferents from color plot data but it is not a fully automated method for adenosine transient finding. Moreover, since the shape of adenosines CV changes with time PCR has difficulties discriminating the secondary peak. When dealing with large amounts of data a completely automated method, which accurately finds spontaneous adenosine transients is necessary.

A

DA + pH
PCs

Only DA
PCs

B

D

C

E
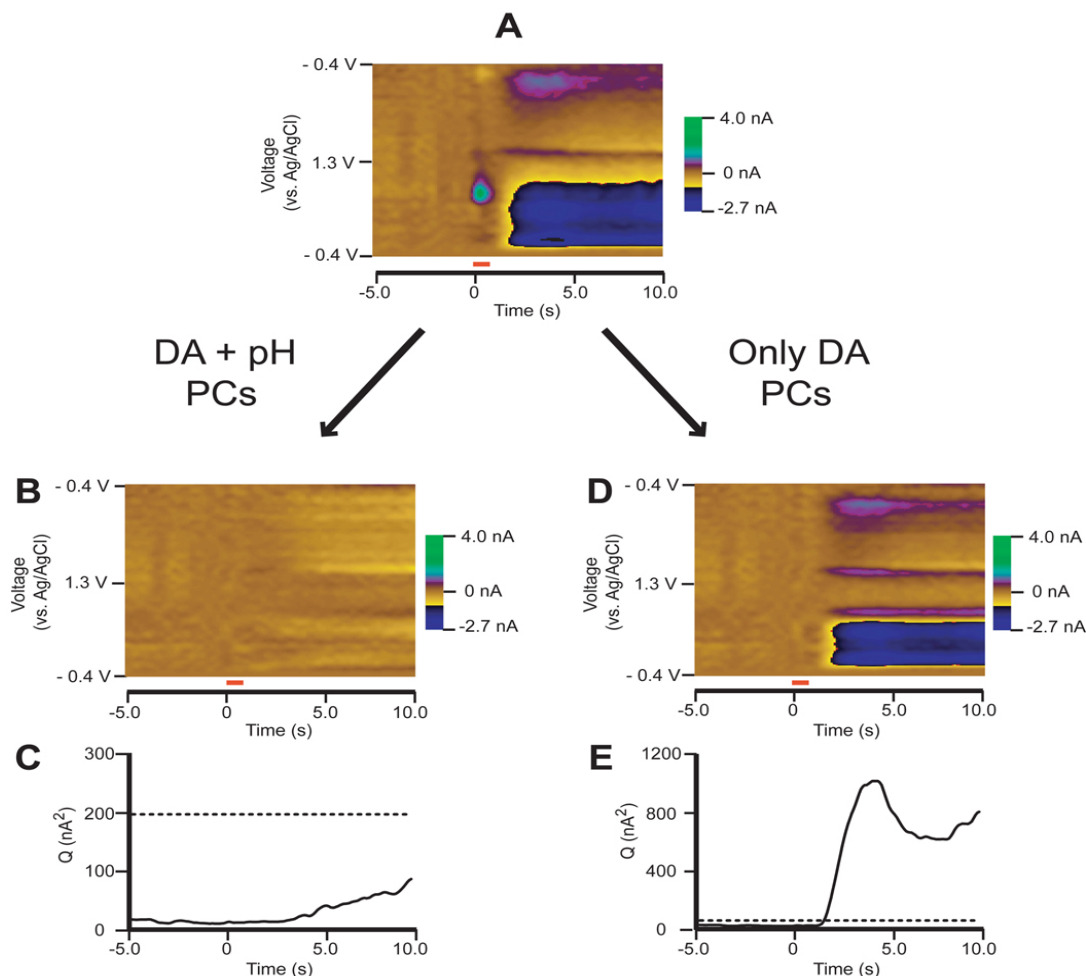
Figure 1.5: Residual plots from *in vivo* stimulation of dopamine in freely moving rats. (A) *in vivo* color plot. (B) Residual plot when dopamine and pH are in training set. (C) Q trace of residual plots where the dashed line represents Q_α at the 95% significance. All Q scores are below this threshold, therefore the training set accurately describes all relevent data. (D) Residual plot with only dopamine in training set. (E) pH is above the Q_α threshold and is retained in the Q trace. Figure taken from Keithley *et al* [27]

1.3.2 Other peak identification techniques

The ability to automate identification of spontaneous adenosine transients is advantageous for two reasons: (1) experimental throughput, since data analysis requires considerable labor (2) accuracy of data analysis within data sets and between researchers. Principal components regression is useful in accurately determining concentrations and using residuals to remove non-relevant data from color plots but is

not completely automated. Other groups have developed methods for automating identification of peaks in chromatography but normally use retention times in analyte recognition[28,29]. In one study retention times were expanded and contracted to fit a target chromatogram. Then the Pearson correlation coefficient was used to determine the degree in which the target and test chromatogram were linearly related[30]. However, spontaneous adenosine release does not produce consistent time markers that enable this type of automation. Furthermore, data from *in vivo* FSCV contains more noise and baseline drift than chromatographic techniques. Finally, since the brain is a complex matrix, unexpected analytes would further muddle these types of automated identification. In this thesis, I describe a computer program that automatically identifies and characterizes adenosine transients, which minimizes labor for data analysis and maximizes accuracy of resulting data.

1.3.3 Automatic identification of adenosine features

In this thesis a straightforward algorithm was designed to identify and characterize random adenosine transients from FSCV color plots. Typically, analysts tabulate adenosine transient features and this process is very tedious and time consuming, often taking 10-18 hours per experiment for a skilled analyst. In a single animal experiment 700 transients can be found and characterized and 24 experiments are needed in order to be statistically relevant and this amounts to 14,400 transients per publication. If an *in vivo* researcher performs 3 experiments per week the data analysis time will be approximately 30-54 hours, which is a significant amount of time. The developed adenosine transient program automatically reads, analyzes, and creates a report of adenosine features in 40 minutes per *in vivo* animal experiment. This thesis demonstrates the reliability of the automated identification program in determining adenosine features and speed of compiling these features. Automatically identifying

analytes in FSCV data will allow researchers to analyze adenosine and is customizable to study other electroactive molecules like dopamine, since the algorithm is modular in design, with fast non-bias computer processing. In conclusion, this thesis will describe a new method for automatically identifying spontaneous adenosine transients, with minimal analyst input, and show how an algorithm can be developed for neurotransmitter identification in *in vivo* FSCV analysis.

1.4 References

(1) Cunha, R. A. (2001) Adenosine as a neuromodulator and as a homeostatic regulator in the nervous system: Different roles, different sources and different receptors. *Neurochem. Int. 38*, 107–125.

(2) Latini, S., and Pedata, F. (2001) Adenosine in the central nervous system: Release mechanisms and extracellular concentrations. *J. Neurochem. 79*, 463–484.

(3) Pajski, M. L., and Venton, B. J. (2013) The mechanism of electrically stimulated adenosine release varies by brain region. *Purinergic Signal. 9*, 167–174.

(4) Mitchell, J. B., Lupica, C. R., and Dunwiddie, T. V. (1993) Activity-dependent release of endogenous adenosine modulates synaptic responses in the rat hippocampus. *J. Neurosci. 13*, 3439–3447.

(5) Nguyen, M. D., Lee, S. T., Ross, A. E., Ryals, M., Choudhry, V. I., and Venton, B. J. (2014) Characterization of Spontaneous, Transient Adenosine Release in the Caudate-Putamen and Prefrontal Cortex. *PLoS One* (Fisone, G., Ed.) *9*, e87165.

(6) Cechova, S., and Venton, B. J. (2008) Transient adenosine efflux in the rat caudate-putamen. *J. Neurochem. 105*, 1253–1263.

(7) Nguyen, M. D., and Venton, B. J. (2015) Fast-scan Cyclic Voltammetry for the Characterization of Rapid Adenosine Release. *Comput. Struct. Biotechnol. J. 13*, 47–54.

(8) Ross, A. E., and Venton, B. J. (2015) Adenosine transiently modulates stimulated dopamine release in the caudate-putamen via A1 receptors. *J. Neurochem. 132*, 51–60.

(9) Swamy, B. E. K., and Venton, B. J. (2007) Subsecond detection of physiological adenosine concentrations using fast-scan cyclic voltammetry. *Anal. Chem. 79*, 744–750.

(10) Fredholm, B. B., AP, I. J., Jacobson, K. A., Klotz, K. N., and Linden, J. (2001) International Union of Pharmacology. XXV. Nomenclature and classification of adenosine receptors. *Pharmacol Rev 53*, 527–552.

(11) Fredholm, B. B., Abbracchio, M. P., Burnstock, G., Daly, J. W., Harden, T. K., Jacobson, K. A., Leff, P., and Williams, M. (1994) Nomenclature and classification of purinoceptors. *Pharmacol. Rev. 46*, 143–56.

(12) Bjorness, T. E., and Greene, R. W. (2009) Adenosine and sleep. *Curr. Neuropharmacol. 7*, 238–45.

(13) Spyer, K. M., and Thomas, T. (2000) A role for adenosine in modulating cardio-respiratory responses: a mini-review. *Brain Res. Bull. 53*, 121–4.

(14) Drury, A. N., and Szent-Györgyi, A. (1929) The physiological activity of adenine compounds with especial reference to their action upon the mammalian heart1. *J. Physiol. 68*, 213–237.

(15) Ohta, A., and Sitkovsky, M. (2001) Role of G-protein-coupled adenosine receptors in downregulation of inflammation and protection from tissue damage. *Nature 414*, 916–920.

(16) Sweeney, M. I. (1997) Neuroprotective effects of adenosine in cerebral ischemia: window of opportunity. *Neurosci. Biobehav. Rev. 21*, 207–217.

(17) Pajski, M. L., and Venton, B. J. (2010) Adenosine release evoked by short electrical stimulations in striatal brain slices is primarily activity dependent. *ACS Chem. Neurosci. 1*, 775–787.

(18) Sharma, R., Engemann, S. C., Sahota, P., and Thakkar, M. M. (2010) Effects of ethanol on extracellular levels of adenosine in the basal forebrain: An in vivo microdialysis study in freely behaving rats. *Alcohol. Clin. Exp. Res. 34*, 813–818.

(19) Grabb, M. C., Sciotti, V. M., Gidday, J. M., Cohen, S. A., and Van Wylen, D. G. L. (1998) Neurochemical and morphological responses to acutely and chronically implanted brain microdialysis probes. *J. Neurosci. Methods 82*, 25–34.

(20) Llaudet, E., Botting, N. P., Crayston, J. A., and Dale, N. (2003) A three-enzyme microelectrode sensor for detecting purine release from central nervous system. *Biosens. Bioelectron. 18*, 43–52.

(21) Kuhr, W. G., and Wightman, R. M. (1986) Real-time measurement of dopamine release in rat brain. *Brain Res. 381*, 168–171.

(22) Baur, J. E., Kristensen, E. W., May, L. J., Wiedemann, D. J., and Wightman, R. M. (1988) Fast-scan voltammetry of biogenic amines. *Anal. Chem. 60*, 1268–1272.

(23) Dryhurst, G. (1977) Purines, in *Electrochemistry of Biological Molecules*, pp 71–185. Elsevier.

(24) Dryhurst, G., and Elving, P. J. (1968) Electrochemical Oxidation of Adenine - Reaction Products and Mechanisms. *J. Electrochem. Soc. 115*, 1014–&.

(25) Johnson, J. A., Rodeberg, N. T., and Wightman, R. M. (2016) Failure of Standard Training Sets in the Analysis of Fast-Scan Cyclic Voltammetry Data. *ACS Chem. Neurosci.* acschemneuro.5b00302.

(26) Jackson, J. E., and Mudholkar, G. S. (1979) Control Procedures for Residuals Associated with Principal Component Analysis. *Technometrics 21*, 341.

(27) Keithley, R. B., Mark Wightman, R., and Heien, M. L. (2009) Multivariate concentration determination using principal component regression with residual analysis. *TrAC - Trends Anal. Chem. 28*, 1127–1136.

(28) Shackman, J. G., Watson, C. J., and Kennedy, R. T. (2004) High-throughput automated post-processing of separation data. *J. Chromatogr. A 1040*, 273–282.

(29) Dixon, S. J., Brereton, R. G., Soini, H. A., Novotny, M. V., and Penn, D. J. (2006) An automated method for peak detection and matching in large gas chromatography-mass spectrometry data sets. *J. Chemom. 20*, 325–340.

(30) Johnson, K. J., Wright, B. W., Jarman, K. H., and Synovec, R. E. (2003) High-speed peak matching algorithm for retention time alignment of gas chromatographic data for chemometric analysis. *J. Chromatogr. A 996*, 141–155.

Chapter 2: Building an algorithm to automatically identify spontaneous adenosine transients

2.1 Introduction

Adenosine is a byproduct of ATP catabolism and important biological nucleoside involved in cell signaling[1], neuromodulation[1,2], and neuroprotection[1,2]. In the brain, adenosine regulates cerebral blood flow and modulates neurotransmission[3,4]. The ability to measure adenosine is important in understanding the roles it plays in neuromodulation and homeostatic regulation. Recently, direct measurements of spontaneous transient adenosine release *in vivo* have been made by fast-scan cyclic voltammetry (FSCV)[5]. These events are spontaneous, rather than stimulated, and last only a few seconds. Several hundred transients can occur in the four hour data collection typical of an *in vivo* experiment. Current data analysis requires a human to pick the transients by hand; a human experienced in analyzing the data can identify a transient approximately every 1.5 minutes. Therefore, if a data set contains 700 transients, then it would take about 18 hours to analyze. Adenosine transient events are seemingly random and do not follow any redily identifible pattern so all the data must be painstakingly analyzed. In addition to being slow, identification by an analyst could be potentially biased. Automating identification of adenosine transients would save time and normalize data analysis between researchers.

Algorithms have been developed to automate identification of molecules in chemical data[6–16]. For *in vivo* electrochemical data, peak identification has used the cyclic voltammogram (CV) as a chemical fingerprint to identify which peak is detected. However, there are 144,000 CVs collected in a four hour voltammetry experiment so they cannot be individually examined. Principal components regression (PCR) uses those cyclic voltammograms to identify compounds in mixture and remove noise from

the data[17]. In particular, PCR has proven to be a powerful tool to separate dopamine from pH shifts.   This method was used previously to identify adenosine and create concentration vs time traces that are analyzed by an analyst to identify adenosine transients.   The problem with PCR for adenosine is that the cyclic voltammogram of adenosine changes over time, with a primary peak that is large in the first few cyclic voltammograms and a secondary peak that grows in over time. Thus, it is hard to select a representative training set, the residuals (i.e. noise) are large, and the residual noise (Q) is often above Q_α, denoting the training is not sufficient to predict the concentration of the neurochemical. The other major problem for finding adenosine transients is that they are random events, with no unique time markers. While many dopamine events are linked to behaviors or cues, finding adenosine transients requires an algorithm that does not use time as a rule for identification.

In this study an algorithm was designed to identify and characterize random adenosine transients from FSCV data. Our program automatically reads, analyzes, and creates a report characterizing the duration, concentration, and event time for each adenosine transient. This automated analysis takes only about 40 minutes to analyze an *in vivo* data set. The program was validated with 4 data sets from 3 independent researchers and compared to the results of human analysts. The program resulting in 10 ± 4% false negatives (FN), due to multiple peaks and high thresholds, and 9 ± 3% false positives (FP), due to random noise that occurs in biological experiments. The algorithm was tested against ATP, histamine, hydrogen peroxide, and pH, known interferents in the brain, and only generated one false positive in 82 measurements. This study demonstrates the reliability of the automated identification program for adenosine and the program is customizable to study other electroactive analytes in the

future. Automated analysis of FSCV data will allow faster data analysis and less analyst bias for identifying and characterizing adenosine *in vivo*.

2.2. Methods

2.2.1 FSCV Transient

The adenosine feature detection algorithm, FSCV Transient, was written in Matlab 2014b (The MathWorks Inc, Natick, MA, USA). First, non-background subtracted and non-filtered FSCV color plot data were exported from High Definition Cyclic Voltammetry (HDCV), a program developed in the Wightman lab [18]. A typical *in vivo* FSCV experiment has files with 80–180s of data and the FSCV transient program individually reads each file for analysis.   After the files are exported an analyst defines three user inputs: maximum (1) primary and (2) secondary oxidation voltage, p_{max} and s_{max}, respectively, and a background subtraction (3) increment value. Then FSCV data is read into the program and convoluted with a 2-D Gaussian filter (size=7, σ=7), which is a low-pass blurring filter. If the algorithm finds adenosine events, then features including event time, concentration, and duration are written to a comma separated value file (*Figure 2.1*, Step 7) until all files are read and analyzed. The program was run on a 3.4 GHz PC computer with Windows 10 for data analysis.

2.2.2 Incremental background subtraction

In the first part of the algorithm (*Figure 2.1*, Steps 1-3), incremental background subtraction is performed by choosing several times for the background and then subtracting the background charging current. The analyst sets the peak voltages for the primary and secondary peaks of adenosine, then *i vs t* data is searched for peaks. For example, an initial background subtraction occurs at t = 1.0 s and then the two *i vs t*

traces are scanned for adenosine peaks (Step 2). Every peak above a set threshold is compiled and concatenated until the end of file. Because you cannot know *a prior* if the chosen time for background subtraction is during a peak and background drift occurs on the order of 90 s, it is best to choose several times and perform background subtraction in order to identify all possible peaks. The program has an analyst-defined increment value, so an increment value of 10.0 s would result in 17 incremented background subtractions in a 180 s file. After all possible adenosine peaks are amassed and duplicates are removed, the algorithm imposes a constraint that a peak must be detected at both the primary and secondary adenosine peak voltages and that the primary adenosine peak maximum has to occur before the secondary peak maximum. The final set of event times identified during incremental background subtraction is subsequently used as seeds during the second part of the program, adjacent background subtraction.
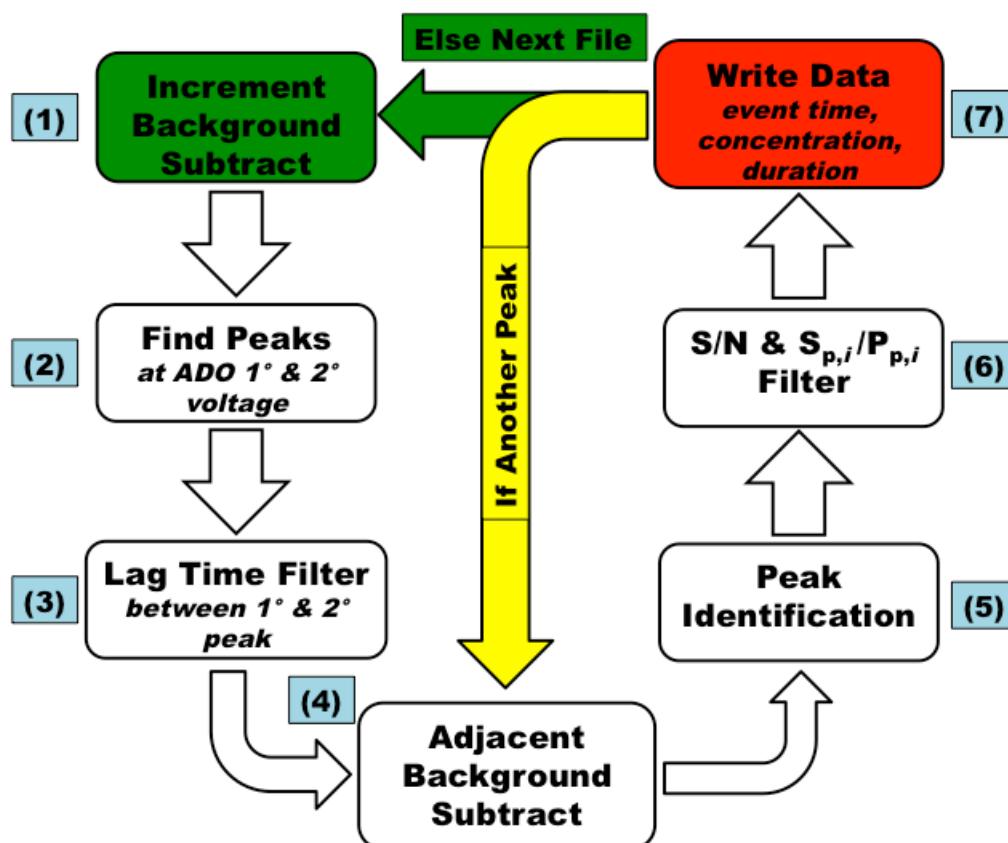
Figure 2.1: Algorithm for FSCV transient. Files are read into the program for incremental background subtracted (1). Analyst defined primary and secondary oxidation voltages are scanned for adenosine peaks (2). Lag time filter is applied to detected peaks to remove spurious peaks (3) and resultant peaks are background subtracted adjacent to the peak (4) then identified similarly to steps 2 and 3 (5). Signal-to-noise and ratio filter are applied to detected peaks to remove spurious peaks (6). Event time, concentration, and duration are written to file until all peaks are investigated (7).

2.2.3 Adjacent background subtraction

During the first part of the algorithm increment background subtraction is performed, which background subtracts the FSCV file at evenly spaced time steps. In the second part of the algorithm (Steps 4-7), adjacent background subtraction is accomplished by performing background subtraction approximately 10 s before, i.e. adjacent, to each individual peak found during incremental background subtraction. Adjacent background subtraction is standard procedure in FSCV because of baseline drift and leads to more accurate measures of adenosine peak characteristics. Some peaks found during the first part of the program are spurious and are rejected as

adenosine during adjacent background subtraction because the two peaks are not identified or the primary peak doesn't precede the secondary peak.

2.2.4 Spurious peak filtering (in vivo)

Peaks are filtered in Step 5 (Step 5 is similar to Step 2 and 3) at analyst-defined thresholds for concentration, duration, and prominence. The minimum concentration that can be detected is usually around 40 nM and the minimum duration for an adenosine transient is 1.0 s. Prominence is the minimum peak height between two consecutive, possibly overlapping peaks. Thresholds are determined from running 5 FSCV data sets in the program and finding the minimum value for concentration that minimizes false negatives and positives. Duration and prominence thresholds are constant and are determined in the same way as concentration but do not need to be adjusted per experiment. Additionally, a signal-to-noise filter is applied to the data and only peaks that have a S/N > 3 are kept, with the noise defined as the SD of the baseline taken adjacent to the peak. Finally, another filter is applied which compares the ratio of the secondary peak max current to primary peak max current, $S_{p,i}$ / $P_{p,i}$ with an empirically determined value for adenosine. The minimum secondary to primary peak ratio for adenosine measured *in vivo* is 0.49, which was determined empirically from 100 *in vivo* adenosine transients. Thus, any peak with a ratio below the threshold of 0.49 is rejected as an adenosine peak. If peaks pass the S/N and ratio filter, they are accepted as adenosine peaks. Programmatic details of background subtraction, data filtering, threshold setting, and peak finding can be found in the *Appendix* at the end of this thesis.

2.2.5 Chemicals

The chemicals used to make phosphate buffered saline (PBS) were all purchased from Fisher Scientific (Fair Lawn, NJ, USA) unless otherwise stated. PBS

buffer was used to test interferents using a flow-injection system[19] and contained (in mM) 131.25 NaCl, 3.0 KCL, 10.0 NaH_2PO_4, 1.2 $MgCl_2$, 2.0 Na_2SO_4, and 1.2 $CaCl_2$. Calcium chloride was purchased from Sigma Aldrich (St. Louis, MO, USA). All aqueous solutions were prepared with deionized water (Milli-Q Biocel; Millipore, Billerica, MA, USA). Adenosine, histamine, and adenosine triphosphate were purchased from Sigma Aldrich and hydrogen peroxide was purchased from Macron (Center Valley, PA, USA). The interferent pH was tested by adjusting pH=7.4 PBS buffer to pH=7.3 or pH=7.5.

2.2.6 Carbon-fiber microelectrodes and FSCV

Carbon-fiber microelectrodes (CMFEs) were prepared with standard fabrication techniques[20]. A single 7 µm T-650 carbon-fiber was aspirated (Cytec Engineering Materials, West Patterson, NJ, USA) into a glass capillary (1.2 mm × 0.68mm; A-M Systems, Inc.,Seqium, WA, USA) which was pulled by a vertical puller (model PE-21; Narishige, Tokyo, Japan) into two microelectrodes. The extended fiber was cut to between 50-150 µm for all data sets. For data sets S1 the interface between the glass and fiber was sealed with epoxy (Epon resin 828; Miller-Stephenson Chemical Co. Inc.; Danbury, CT, USA) and 14% wt. m-phenylenediamine hardener (Acros Organics, Morris Plains, NJ, USA). For Data sets S2, S3, and S4, in Table 2.1, and all flow-injection experiments the fibers were not epoxied.

Fast-scan cyclic voltammetry (FSCV) was used to monitor electroactive species in animal and flow-injection experiments. The waveform and data collection was computer controlled by High Definition Cyclic Voltammetry (HDCV) (gift of Mark Wightman, UNC at Chapel Hill)[18]. A Dagan ChemClamp potentiostat (Dagan Corporation; Minneapolis, MN, USA) was used to apply voltage to the CFME. All electrodes were scanned from a holding potential of -0.40 V and scanned to a switching

potential of 1.45 V and back at 10 Hz versus a Ag/AgCl reference electrode, at a scan rate of 400 V/s. All data was background subtracted to remove any non-Faradic currents by taking the mean of 10 CVs and background subtracting that vector from the data set. All *in vitro* interferent tests were performed using flow-injection analysis by comparing 1.0 µM adenosine to 1.0 µM interferent in PBS buffer.

2.2.7 Data sets analyzed

In vivo data set (S1) was measured in the caudate putamen and data sets (S2 and S3) were measured in the hippocampus according to procedures previously described[5]. The brain slice data was measured in the prefrontal cortex according to procedures previously described[21].

2.2.8 Error analysis

Sensitivity, precision, and accuracy were calculated from true positive (TP), false positive (FP), and false negative (FN) values determined from analyst validation of FSCV transient algorithm results (Results and discussion *3.2.3.*). Sensitivity or recall is the fraction of relevant peaks that are returned by the algorithm from the data set.

$$\frac{TP}{TP+FN} \quad (1)$$

Precision or positive predictive value is the fraction of peaks returned by the algorithm from the data set that are relevant peaks.

$$\frac{TP}{TP+FP} \quad (2)$$

Accuracy was calculated from the F_1 score, which is the harmonic mean of sensitivity and precision. The harmonic mean weights sensitivity and precision equally.

$$2 \frac{sensitivity \times precision}{sensitivity + precision} \quad (3)$$

The F_1 score was calculated because the amount of true negatives, or the number of times the algorithm missed a spurious peak, is unable to be calculated. Values calculated from equations (1-3) are between 0 and 1.0, with 1.0 being maximum sensitivity, precision, and accuracy for analyst validation. Data are presented as mean ± standard deviation.

2.3. Results and discussion

2.3.1 Adenosine feature detection (algorithm)

To publish a paper for an *in vivo* experiment many animal experiments are needed and analyzing the resulting data sets takes much more time than collecting the data. Typically, researchers collect 4 hours of data from a single animal experiment but spend 10-18 hours identifying transients and calculating the event time, concentration and duration of the transients by hand. With FCSV cyclic voltammograms are taken 10 times per second, and multiple CVs are often viewed as a color plot, with data stacked as a function of time. The resulting color plot (*Figure 2.2B*) forms a three-dimensional plot of voltage and current as a function of time. Data files are typically 180s worth of data, 20-80 files per experiment. Thus, 36,000-144,000 individual cyclic voltammograms are collected and must be analyzed per experiment. Thus, an automated, unbiased method of analyzing thousands of cyclic voltammograms to identify hundreds of adenosine transients is needed. Building a computer program will automate this process, normalize data analysis between researchers and allow more time to conduct experiments and interpret experimental results.
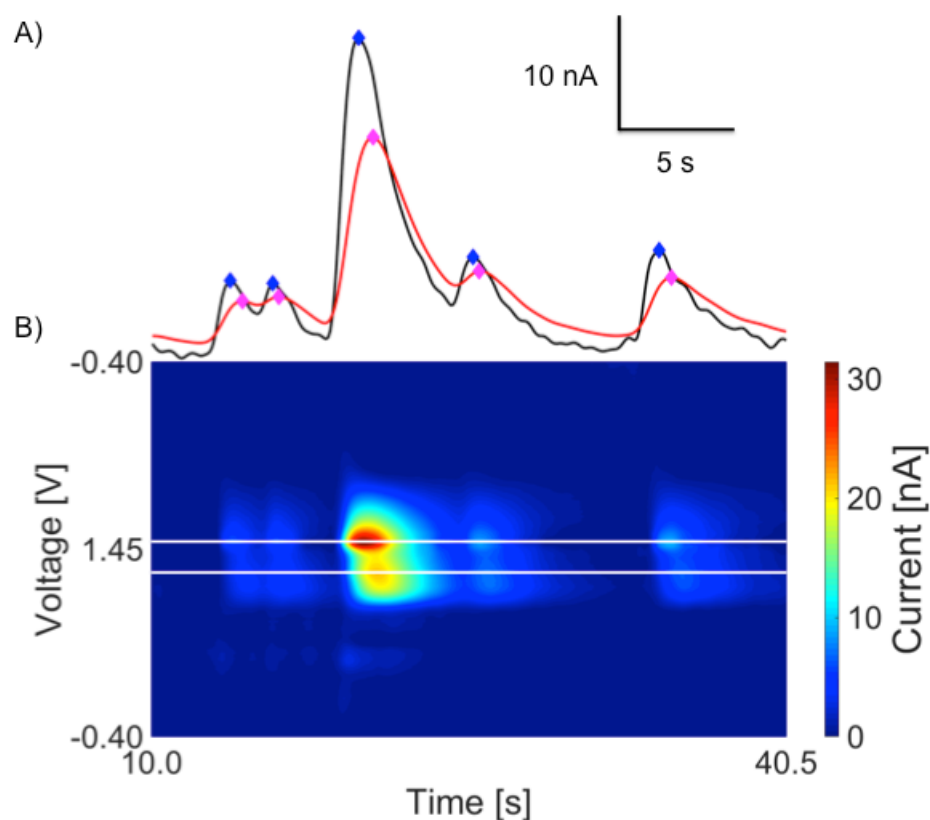
Figure 2.2: *in vivo* spontaneous adenosine transients from the hippocampus brain region. A) *i vs t* trace with 5 transients detected. The primary peak maximums (black trace) occur before secondary peak maximums (red trace) and is the basis for the lag time filter. The $S_{p,i}$ / $P_{p,i}$ ratio is > 0.50, thus this peaks would be accepted as adenosine transients. B) False color plot of adenosine transients with white lines representing the primary and secondary oxidation voltages.

In the brain, spontaneous adenosine release is a random process so the time they will occur cannot be predicted. Electrochemical properties of adenosine must be exploited to identify adenosine transients. Adenosine is irreversibly oxidized[22] around 1.4 V and forms a primary product, which is subsequently irreversibly oxidized around 1.0 V to form a secondary product in FSCV experiments[4]. The peak oxidation voltage for both products is constant at a singular electrode and therefore these peaks are used to identify adenosine. The primary product is the precursor for the secondary product so there is a lag time between peak maximum, which is seen in current versus time traces in *Figure 2.2A*[23]. The peak maximums are marked as diamonds and the primary peak

always occurs before the secondary. Additionally, as seen in the false color plot in *Figure 2.2B,* there is lag time between primary and secondary peak formation with the primary occurring before the secondary.

Lag time and peak oxidation voltage provides enough information for the algorithm to successfully identify adenosine transients. Primary and secondary max voltages, p_{max} and s_{max}, respectively, are analyst-defined and *i vs t* traces at these voltages are scanned for peaks above a threshold, illustrated by the white lines in the color plot in *Figure 2.2B*. For a peak to be identified as adenosine it must have a peak both at the primary and secondary voltages and the secondary peak must lag the primary peak by at least 0.1 s but be within 2.5 s of the primary peak. Additionally, another filter is applied which compares the ratio of adenosines secondary peak max current to primary peak max current, $S_{p,i}$ / $P_{p,i}$. If peaks pass the threshold, lag time criteria, and secondary-to-primary peak ratio, then the signal-to-noise ratio (SNR) is calculated for each adenosine peak. If the resultant peak is above 3 × SD of baseline noise, the program saves the event time, peak concentration, duration, and SNR. This identification algorithm is the basis for automatically identifying and tabulating spontaneous adenosine transients, in large data sets with many peaks.

2.3.2 Background Subtraction

In FSCV experiments, stable non-faradaic currents occur due to background charging of the surface of the CFME and these background currents are subtracted to study Faradaic redox reactions. Because the location of adenosine transients is not known *a priori*, the program first picks several places for background subtraction, at defined increments, and then background subtracts each data set at these times. By examining the same data set background subtracted from different places, we correct for

2 fundamental problems. First, if we only picked one background subtraction time, it may inadvertently be during an adenosine peak and the results would not be interpretable. Second, the background current drifts over time so background subtraction should be performed as near to a peak as possible to accurately identify and define all possible adenosine peak characteristics. Practically, the algorithm accomplishes this by first reading in a non-background subtracted data file into the program. Next, the program background subtracts at the start of the file, completely scans p_{max} and s_{max} for peaks, and then uses the identification algorithm to determine if any adenosine transients are present in the data set. After the initial background subtraction, the program iterates at the analyst-defined increment value (usually 10 s) (*Figure 2.3A-C*) repeating the procedure of doing a background subtraction and identifying adenosine transients using that background file. This iteration continues at the defined times until the end of the file. All potential adenosine transients are identified by event time and are saved for later use. Since the program is deterministically incrementing and adenosine transients are random, the program will detect spurious peaks, which need to be further explored. The purpose of this part of the identification algorithm is to tabulate all possible adenosine transient times for each color plot file.

During the incrementing part of the program the algorithm casts a wide net in order to obtain all possible adenosine events. This strategy works in gathering all peaks, real and spurious, but in order to determine if a peak is adenosine, it more standard to do background subtraction directly adjacent to the peak[5]. The transient event times, obtained in the first part of the program, are used in the second part of the algorithm
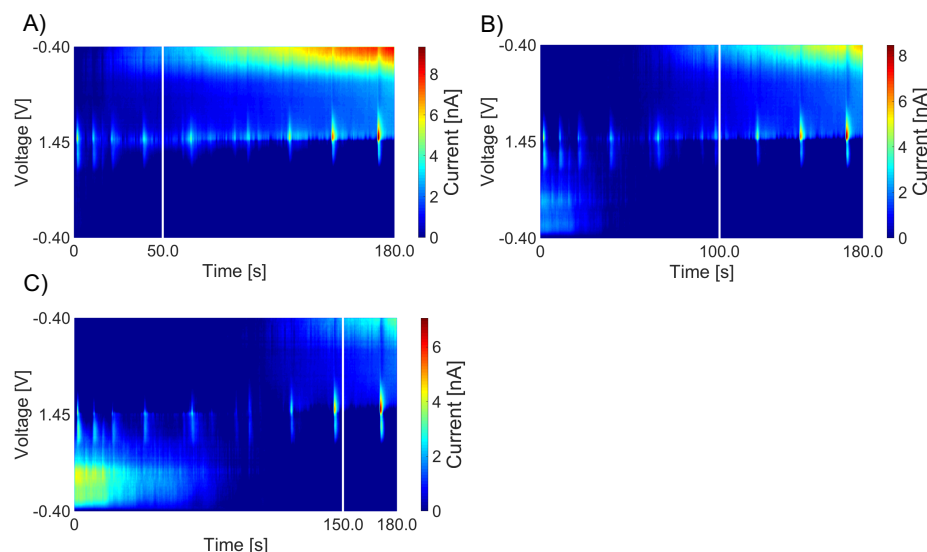
A)

B)

C)

Figure 2.3: Incremental background subtraction of *in vivo* adenosine transients from the hippocampus brain region. A-C) Incrementally background subtracted FSCV data every 50 seconds. The white vertical lines in the files show when the background was taken.

where each location is background subtracted adjacent to the peak. Adjacent background subtraction is performed approximately 2.5 s before each peak. By subtracting the background adjacent to a presumed adenosine peak, the concentration change and duration of adenosine can be more accurately determined. If spurious peaks are collected during the increment part of the algorithm, they are often rejected when adjacent background subtraction is performed. Thresholds, in both the increment and adjacent background subtraction are set according to experimental conditions. The incrementing part of the program has wide scope for peak detection and threshold values are set lower to collect all potential transients. However, during the adjacent background subtraction, threshold values should be higher and therefore more discriminating. The goal of adjacent background subtraction is to accurately probe each probable peak location and positively identify adenosine transients by removing any spurious noise peaks.

2.3.3 Analyst validation of adenosine transient program

An additional way of determining the robustness of the adenosine transient program is to run the data sets through the programs algorithm and validate the output data by an analyst. This comparison helps to verify the algorithm success at identifying adenosine transients, and test how the algorithm performs compared to analysts. The goal is to minimize the amount of false negatives (FN) and positives (FP) obtained from the adenosine algorithm. Using the event time output from the program, an analyst verified each peak identified as adenosine with High Definition Cyclic Voltammetry (HDCV), a program developed in the Wightman lab[18]. The results are tabulated in *Table 2.1*. If the algorithm selects a peak that was not identified as adenosine by the analyst, it is counted as a FP. Moreover, if the analyst determines that the program missed an adenosine peak it is counted as FN. Each data set was measured in an independent animal experiment and the data sets were obtained from three independent experimenters. The first data set (S1)*,* an *in vivo* measurement in the caudate putamen, an analyst determined 41 adenosine transients and the algorithm resulted in 5% FN and 2% FP. In data set S2, an *in vivo* measurement in the hippocampus, an analyst determined 397 adenosine transients and the algorithm resulted in 10% FN and 10% FP. In data set S3, another *in vivo* measurement in the hippocampus, the algorithm resulted in 8% FN and 9% FP. Finally, in data set S4, a brain slice experiment from the prefrontal cortex, the algorithm resulted in 16% FN and 7% FP. The reason for S4 having a higher FN percentage than the other sets is due to thresholds being adjusted to minimize the selection of FP. Overall, analyst validation of the adenosine algorithm resulted in $10 \pm 4\%$ FN and $9 \pm 3\%$ FP in a total of 640 confirmed adenosine transients. These results suggest that the adenosine algorithm is able to discern adenosine, with a high degree of certainty, from noise that is present during animal experiments.

Table 2.1: Analyst validation of adenosine transient data sets. S1 is an *in vivo* measurement in the caudate putamen. S2 and S3 are *in vivo* measurements in the hippocampus. S4 is a brain slice experiment from the prefrontal cortex. Each data set is an independent experiment

	S1	S2	S3	S4	Total
False Negative	5%	10%	8%	16%	10 ± 4%
False Positive	2%	10%	9%	7%	9 ± 3%
n	41	397	112	90	640

Any automated method for measurement and identification is subject to false negatives and false positives. Since many common interferents found in the brain have been rejected by flow-injection analysis experiments (*vide infra*), FP are mainly generated from random noise in the data. The main reason the adenosine transient algorithm will generate FN is due to multiple peak maximums occurring in a single peak. False negatives due to multiple peaks can be corrected for in the program by adjusting the prominence threshold, which is the minimum peak height between two consecutive, possibly overlapping peaks. However, it will always be difficult to measure the concentration and duration of multiple peaks that do not go back to baseline in between. Alternatively, the transient program will reject peaks if the data is noisy and the threshold is set above the level of small adenosine transients. If thresholds are high during the incremental background subtraction, adenosine transients are rejected and cannot be measured during adjacent background subtraction. One strategy for setting up thresholds in both the incremental and adjacent background subtraction is to minimize the amount of FPs but this will ultimately increase FN, as seen in the brain slice experiment. Alternatively, setting thresholds properly in both background subtraction parts of the adenosine algorithm can achieve the minimization of both FP and FN. Overall, analyst validation of adenosine transient program had a mean precision of 0.91 ± 0.01, sensitivity of 0.90 ± 0.04, and accuracy of 0.90 ± 0.02. An accuracy of 0.90 is sufficient for the FSCV transient algorithm because analysts also fail to detect adenosine transients in data when counting.

2.3.4 Testing biologically relevant interferents (in vitro)

The brain is a complex organ with multiple electroactive molecules that could interfere with adenosine detection. During adjacent background subtraction, the algorithm checks if peaks exist at p_{max} and s_{max} voltages and that a lag time exists between these peaks. In order to test the robustness of the algorithm, adenosine and possible interferents were measured in a flow cell. Adenosine was measured in a flow-injection system to determine p_{max} and s_{max}, the voltages for the primary and secondary peaks. Since the oxidation voltage of adenosine remains constant during animal experiments, p_{max} and s_{max} are scanned for potential adenosine peaks. In order for an interferent to be counted as adenosine transient, the interferent must have oxidation potentials near the p_{max} and s_{max} of adenosine, a $s_{p,i}$ /$p_{p,i}$ ratio above adenosines threshold, and importantly the primary peak must occur before the secondary. Adenosines $s_{p,i}$ /$p_{p,i}$ ratio, determined from *in vitro* analysis to be 0.34, was calculated from 20 injections of 1 µM adenosine at 5 electrodes, which was the minimum ratio calculated (mean=0.6±0.2, range=0.3%-1.02%, 20 injections, 5 electrodes). The reason for the large standard deviation and upper range being above 1.0 is due to electrode noise.

To determine if the algorithm generates false positives (FP) (*Figure 2.4*) adenosine triphosphate, histamine, hydrogen peroxide, and pH were tested as possible adenosine interferents. First, 1 µM ATP (*Figure 2.4B*) was tested with the algorithm values for adenosine (*Figure 2.4A*) in order to try to generate FP and no data was omitted from analysis. ATP differs from adenosine by only three phosphate groups and has the same electrochemical moiety[23]. As seen in the *i vs t* trace for ATP, there are primary and secondary peaks. However, the max $s_{p,i}$ /$p_{p,i}$ ratio is 0.27 (mean=0.18±0.05,

range=0.07-0.27, 15 injections, 6 electrodes), which is below the threshold for adenosine

and the secondary occurs before the primary maximum. Thus, ATP fails to be identified
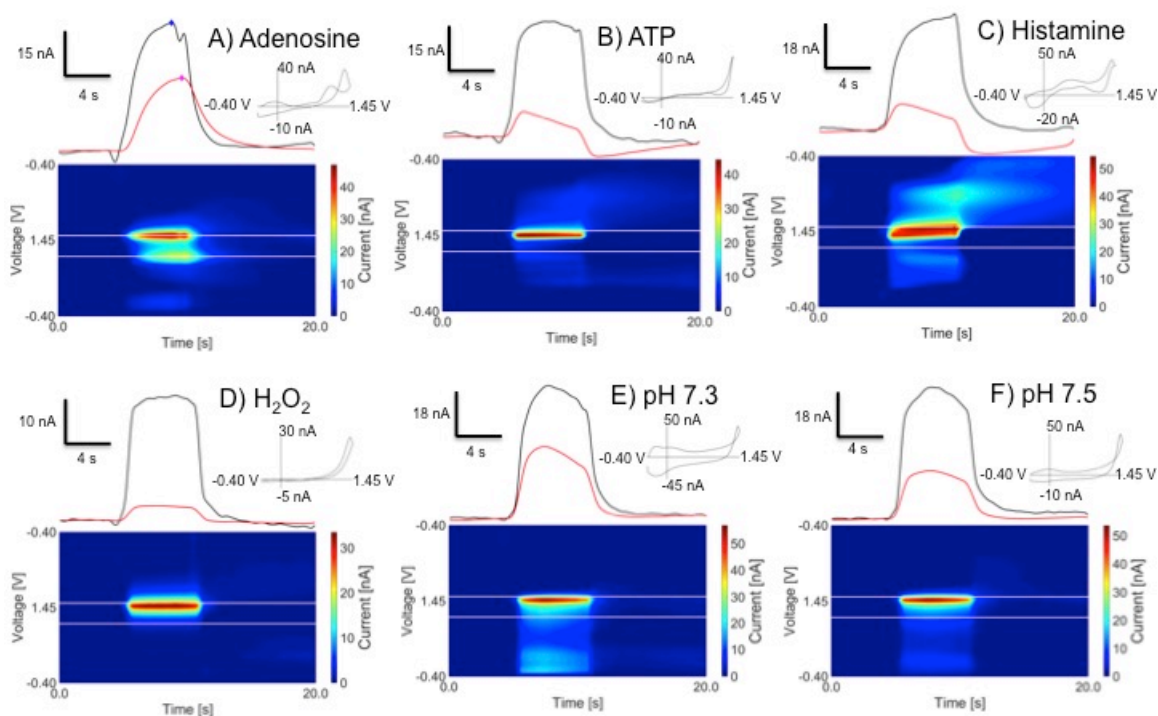

Figure 2.4: *in vitro* testing of biologically relevant interferents. *i vs t* traces, cyclic voltammograms, and false color plots for A) adenosine, B) ATP, C) histamine, D) hydrogen peroxide, E) pH 7.3 shift, F) pH 7.5 shift. Interferents are rejected by the algorithm due to s_{max} occurring before p_{max} and an interferent max $s_{p,i}$ /$p_{p,i}$ ratio below adenosines minimum $s_{p,i}$ /$p_{p,i}$ ratio.

as adenosine. Furthermore, comparing adenosines cyclic voltammogram with ATP's,

adenosine displays a more pronounced secondary peak than ATP. The algorithm also

rejected histamine, a molecule whose cyclic voltammogram is similar to adenosine[24].

The secondary peak for histamine is at 0.76V compared to 1.06V for adenosine. Thus,

the *i vs t* trace for 1 µM histamine (*Figure 2.4C*) displays the s_{max} occurring before p_{max}

and a max $s_{p,i}$ /$p_{p,i}$ ratio of 0.24 (mean=0.15±0.07, range=0.03-0.24, 20 injections, 6

electrodes), which is below the threshold for adenosine of 0.34. Setting p_{max} and s_{max}

constant exploits adenosines intrinsic oxidation potentials. Hydrogen peroxide (*Figure

2.4D*) is another possible interferent of adenosine[21], and is rejected as a transient from

the computer algorithm because the max $s_{p,i}$ /$p_{p,i}$ ratio of 0.18 (mean=0.08±0.05,

range=0.03-0.18, 19 injections, 6 electrodes), is below adenosines threshold and because it has no secondary peak. The $s_{p,i}$ /$p_{p,i}$ ratio can be calculated with no secondary peak from noise present in the data. Finally, pH changes of ±0.1 of pH 7.4 PBS buffer (*Figure 2.4E and 2.4F*) were tested with the algorithm. The max $s_{p,i}$ /$p_{p,i}$ ratio for pH 7.3 and pH 7.5 are below adenosines ratio threshold and have max ratio values of 0.30 (mean=0.19±0.09, range=0.06-0.30, 13 injections, 6 electrodes) and 0.37 (mean=0.20±0.09, range=0.07-0.37, 15 injections, 6 electrodes), respectively. The s_{max} occurs before p_{max} in both pH shifts. Only one sample of pH 7.5 resulted in a $s_{p,i}$ /$p_{p,i}$ ratio of 0.37, which is above adenosines *in vitro* minimum threshold of 0.34, the other three runs on this electrode were below the threshold and therefore rejected. Biologically relevant interferents do not result as false positives for adenosine during *in vitro* experiments, which demonstrates the robustness of the adenosine identification algorithm.

2.3.5 in vivo testing of stimulated histamine

Oxidation voltages in animal experiments can differ from voltages observed during *in vitro* experiments. To further check the robustness of the adenosine algorithm, p_{max} and s_{max} generated from adenosine transients were verified against stimulated histamine data to determine if *in vivo* histamine would be counted as adenosine. Measurements were made in the premammillary nucleus with a stimulating electrode in the medial forebrain bundle region. Histamine has been suspected as a possible interferent in the identification of adenosine due to the similarity in cyclic voltammograms and histamine's secondary oxidation peak formation[24]. The flow-injection experiment demonstrated that p_{max} for both adenosine and histamine are similar but the s_{max} in histamine occurs 0.3 V below the s_{max} of adenosine. In order to test for a FP, the p_{max} and s_{max} obtained from adenosine transients, at 1.41 V and 1.18 V, respectively, were

input into the program and stimulated histamine data was analyzed. The adenosine

identification algorithm did not generate a FP for stimulated histamine (*Figure 2.5*). The

p_{max} of histamine and adenosine are nearly identical but the s_{max} of histamine has a

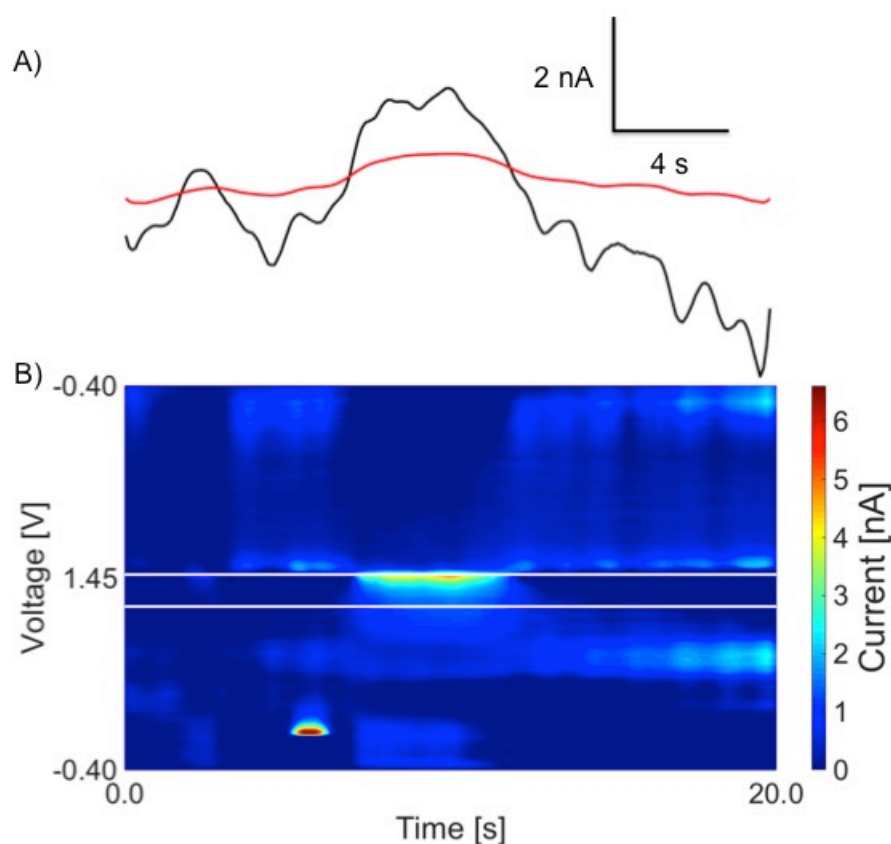difference of 0.07 V from the s_{max} of adenosine. However, the reason

Figure 2.5: *in vivo* stimulated histamine from the premammillary nucleus. A) Primary (black) and secondary (red) oxidation peak *i vs t* traces of stimulated histamine fail lag time filter because maximums occur at the same time. B) False color plot white lines are p_{max} and s_{max} generated from adenosine transients in data set 2 (S2).

that histamine fails the algorithm is due to the primary peak max occurring at the same

time secondary peak max (*Figure 2.5A*); therefore, no time-lag exists between peaks.

Histamines *in vitro* data (*Figure 2.4C*) displays the secondary peak occurring before the

primary, which also fails the lag time filter. Viewing the color plot for stimulated

histamine in *Figure 2.5B*, the secondary peak formation is barely visible and is slightly

below adenosines s_{max} voltage vector. The adenosine algorithm in both in vitro and

animal experiments does not detect histamine and it is concluded that during *in vivo* experiments detected transients are not histamine. This validation shows the algorithm is good at distinguishing histamine as an interferent both during *in vitro* calibration experiments and *in vivo*.

2.4. Conclusion

The ability to automate the identification of adenosine transient features will reduce the hours researchers spend on monotonous data analysis and normalize results between researchers. The first iteration of the algorithm was building a structure to acquire all possible adenosine peaks by incrementing background subtraction. Next, to maximize accuracy of adenosine feature detection adjacent background subtraction was added to the algorithm. Moreover, signal-to-noise, ratio filters, and analyst-defined thresholds can be adjusted to analyze independent data sets from multiple researchers. In summary, this program can save more than two hundred hours of repetitive data analysis per publication. This accumulated time can be used to conduct more experiments and therefore increase laboratory throughput.

2.5. Future directions

The initial development of the algorithm generated promising result for identifying adenosine transients in FSCV data sets. Since CFMEs are manufactured in our lab and have different sensitivities, adjusting thresholds is expected for independent animal experiments. After a CFME is equilibrated in tissue the resulting noise is stable and is relativity constant during animal experiments. One way to automatically calculate threshold values is by calculating $3 \times SD$ of the primary peak noise and set the peak concentration threshold with this value. Determining the concentration threshold is probably the most time consuming step for this algorithm to work and automatically

setting this value would be advantageous.

This software is modular and has the ability to be programmed to identify numerous analytes detected in FSCV data other than adenosine. Moreover, the strategy for detecting adenosine transients could be extended to other analytes like oxygen and dopamine. For instance, oxygen and adenosine release are correlated, so the program could scan the reduction voltage for oxygen in a window after an adenosine peak is detected to scan for oxygen transients[25]. Additionally, the algorithm could scan oxidation and reduction voltages for dopamine, with the time-lag threshold set to zero, and automatically detect spontaneous dopamine transients. To make a dopamine algorithm more robust a reduction to oxidation ratio threshold would be empirically calculated. This ratio would help reduce possible FP from being accepted by the program and make the program more robust. Alternatively, dopamine could be pre-processed by PCR to remove noise from the color plot data and subsequently post-processed by the dopamine transient algorithm. Thus, as long as there is enough information to make rules about detection from the electrochemical data, there are limitless possibilities for this modular spontaneous transient program in analyzing electroactive species. An automated analyte identification algorithm saves hundreds of hours of time in tedious peak feature detection and will normalize data between animals, researchers, and institutions.

2.6. References

(1) Cunha, R. A. (2001) Adenosine as a neuromodulator and as a homeostatic regulator in the nervous system: Different roles, different sources and different receptors. *Neurochem. Int. 38*, 107–125.

(2) Latini, S., and Pedata, F. (2001) Adenosine in the central nervous system: Release mechanisms and extracellular concentrations. *J. Neurochem. 79*, 463–484.

(3) Cechova, S., and Venton, B. J. (2008) Transient adenosine efflux in the rat caudate-putamen. *J. Neurochem. 105*, 1253–1263.

(4) Nguyen, M. D., and Venton, B. J. (2015) Fast-scan Cyclic Voltammetry for the Characterization of Rapid Adenosine Release. *Comput. Struct. Biotechnol. J. 13*, 47–54.

(5) Nguyen, M. D., Lee, S. T., Ross, A. E., Ryals, M., Choudhry, V. I., and Venton, B. J. (2014) Characterization of Spontaneous, Transient Adenosine Release in the Caudate-Putamen and Prefrontal Cortex. *PLoS One* (Fisone, G., Ed.) *9*, e87165.

(6) Dixon, S. J., Brereton, R. G., Soini, H. A., Novotny, M. V., and Penn, D. J. (2006) An automated method for peak detection and matching in large gas chromatography-mass spectrometry data sets. *J. Chemom. 20*, 325–340.

(7) Hastings, C. A., Norton, S. M., and Roy, S. (2002) New algorithms for processing and peak detection in liquid chromatography/mass spectrometry data. *Rapid Commun. Mass Spectrom. 16*, 462–467.

(8) Johnson, K. J., Wright, B. W., Jarman, K. H., and Synovec, R. E. (2003) High-speed peak matching algorithm for retention time alignment of gas chromatographic data for chemometric analysis. *J. Chromatogr. A 996*, 141–155.

(9) Katajamaa, M., and Orešič, M. (2007) Data processing for mass spectrometry-based metabolomics. *J. Chromatogr. A 1158*, 318–328.

(10) Shackman, J. G., Watson, C. J., and Kennedy, R. T. (2004) High-throughput automated post-processing of separation data. *J. Chromatogr. A 1040*, 273–282.

(11) Boiret, M., Gorretta, N., Ginot, Y. M., and Roger, J. M. (2016) An iterative approach for compound detection in an unknown pharmaceutical drug product: Application on Raman microscopy. *J. Pharm. Biomed. Anal. 120*, 342–351.

(12) Fong, S. S., Rearden, P., Kanchagar, C., Sassetti, C., Trevejo, J., and Brereton, R. G. (2011) Automated Peak Detection and Matching Algorithm for Gas Chromatography−Differential Mobility Spectrometry. *Anal. Chem. 83*, 1537–1546.

(13) Ho, T. J., Kuo, C. H., Wang, S. Y., Chen, G. Y., and Tseng, Y. J. (2013) True ion pick (TIPick): A denoising and peak picking algorithm to extract ion signals from liquid chromatography/mass spectrometry data. *J. Mass Spectrom. 48*, 234–242.

(14) Ji, C., Li, S., Reilly, J. P., Radivojac, P., and Tang, H. (2016) XLSearch: a Probabilistic Database Search Algorithm for Identifying Cross-Linked Peptides. *J. Proteome Res. 15*, 1830–41.

(15) Wang, Z., Lin, L., Harnly, J. M., Harrington, P. D. B., and Chen, P. (2014) Computer-aided method for identification of major flavone / flavonol glycosides by high-performance liquid chromatography – diode array detection – tandem mass spectrometry (HPLC – DAD – MS / MS). *Anal Bioanal Chem* 7695–7704.

(16) Woldegebriel, M., and Vivó-Truyols, G. (2015) Probabilistic Model for Untargeted Peak Detection in LC-MS Using Bayesian Statistics. *Anal. Chem. 87*, 7345–55.

(17) Keithley, R. B., Mark Wightman, R., and Heien, M. L. (2009) Multivariate concentration determination using principal component regression with residual analysis. *TrAC - Trends Anal. Chem. 28*, 1127–1136.

(18) Bucher, E. S., Brooks, K., Verber, M. D., Keithley, R. B., Owesson-White, C., Carroll, S., Takmakov, P., McKinney, C. J., and Wightman, R. M. (2013) Flexible software platform for fast-scan cyclic voltammetry data acquisition and analysis. *Anal. Chem. 85*, 10344–10353.

(19) Strand, A. M., and Venton, B. J. (2008) Flame etching enhances the sensitivity of carbon-fiber microelectrodes. *Anal. Chem. 80*, 3708–3715.

(20) Huffman, M. L., and Venton, B. J. (2008) Electrochemical properties of different carbon-fiber microelectrodes using fast-scan cyclic voltammetry. *Electroanalysis 20*, 2422–2428.

(21) Ross, A. E., and Venton, B. J. (2014) Sawhorse waveform voltammetry for selective detection of adenosine, ATP, and hydrogen peroxide. *Anal. Chem. 86*, 7486–7493.

(22) Dryhurst, G. (1977) Purines, in *Electrochemistry of Biological Molecules*, pp 71–185. Elsevier.

(23) Swamy, B. E. K., and Venton, B. J. (2007) Subsecond detection of physiological adenosine concentrations using fast-scan cyclic voltammetry. *Anal. Chem. 79*, 744–750.

(24) Samaranayake, S., Abdalla, A., Robke, R., Wood, K. M., Zeqja, A., and Hashemi, P. (2015) In vivo histamine voltammetry in the mouse premammillary nucleus. *Analyst 140*, 3759–65.

(25) Wang, Y., and Venton, B. J. (2016) Correlation of transient adenosine release and oxygen changes in the caudate-putamen. *J. Neurochem.* 2–4.

Appendix: Programming Details

<u>Code for incremental background subtraction</u>

All code was written in Matlab version 2016a.

A non-background subtracted file **(xyz)** is read into the program. A 10 column vector **(bd)** is taken from this matrix from lower **(ll)** and upper **(ul)** limits, which are incremented by the code.

CODE:
```
bd=xyz(:,ll:ul);
```

The column vector **(bd)** is then averaged at each row **(r)** to create a single vector **(mnbd)**.

CODE:
```
for j=1:r
    mnbd(j,:)=sum(bd(j,:))/(ul-ll);
end
```

Each column **(c)** of the non-background subtracted file **(xyz)** is subtracted by the single vector **(mnbd)** to create a background subtracted file **(d)**.

CODE:
```
for j=1:c
    d(:,j)=xyz(:,j)-mnbd; % d=difference
end
```

The lower and upper limits, **ll** and **ul**, respectively, are deterministically incremented during incremental background subtraction by an analyst defined incrementation value.

<u>Code for adjacent background subtraction</u>

During adjacent background subtraction the "seed" values found during incremental background subtraction **(singleData)** are used to accurately background subtract close to each peak. The same code provided above is used for background subtraction but each background subtraction is performed adjacent to each seed peak. Therefore, the lower **(ll)** and upper **(ul)** limits are more accurately defined by subtracting a **singleData** value by 25 (2.5 seconds at 10 Hz) to determine the upper limit for background subtraction. The lower limit is taken by subtracting the upper limit by 9.

CODE:
```
ul=singleData(i)-25;
ll=ul-9;
```

Code for convoluting background subtracted data

   The data used for adjacent background subtraction is convoluted with a **Gaussian filter** (size **7×7**, σ=**2**), which acts as a low-pass blurring filter. This is performed to blur the 2D data **(M)** and prevent multiple peaks from being detected, due to noisy biological data.

CODE:
```
filt=(fspecial('gaussian',7,2));
C=conv2(single(M),filt,'same') ;
```

Code for smoothing data for peak detection of final data

   No data smoothing was applied to the primary peak for final peak detection.

*Code for peak finding and feature detection (Matlab function **findpeaks**)*

   Matlab has a function that searches a single vector **(primary)** for local maxima. Analyst defined thresholds for minimum peak height **(MinPeakHeight)**, minimum **(MinPeakWidth)** and maximum **(MaxPeakWidth)** peak width with width reference at one-half peak height **(halfheight)**, and minimum peak prominence **(MinPeakProminence)** are set. Thresholds for minimum and maximum peak width are at 10 Hz. Thresholds for minimum peak height are set for each animal experiment and all other thresholds remain constant for all experimental datasets. Function outputs for peak currents **(PKS1)**, event time **(LOCS1)**, duration **(WDTH1)**, and current prominences **(PROM1)**, are determined for each dataset primary max vector.

CODE:
```
[PKS1,LOCS1,WDTH1,PROM1]=findpeaks(primary,'MinPeakHeight',1.0000,...
   'MaxPeakWidth',150,'MinPeakWidth',6,'MinPeakProminence',0.6390,...
   'WidthReference','halfheight');
```

Code for identification of final peaks during adjacent background subtraction

   First, the program takes the **seed** from incremental background subtraction's **singleData**.

CODE:
```
seed=singleData(iteration);
```

Next, the final event time data, acquired from adjacent background subtraction, is stored in **loc1** of **dataFinal**.

CODE:
```
dataFinal=[loc1;loc2;pks1;pks2;prom1;prom2;...
   wdth1;wdth2;sn1z;sn2z;CUtime;FILE];
```

Finally, the event time location **(dataFinal(1,i))** is subtracted from the **seed** value and the resulting difference must be less than 2.2 seconds (data is taken at 10 Hz) of each other.

CODE:
```
seed-dataFinal(1,i) <= 22;
```

The value of 2.2 seconds was used due to multiple peaks being present in the data. This value can be adjusted to suit experimental needs.