DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY

Biological Time Series Analysis Using a Context Free Language: Applicability to Pulsatile Hormone Data

Citation

Dean, Dennis A., Gail K. Adler, David P. Nguyen, and Elizabeth B. Klerman. 2014. "Biological Time Series Analysis Using a Context Free Language: Applicability to Pulsatile Hormone Data." PLoS ONE 9 (9): e104087. doi:10.1371/journal.pone.0104087. http://dx.doi.org/10.1371/journal.pone.0104087.

Published version

https://doi.org/10.1371/journal.pone.0104087

Link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:12987415

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at
https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFlOWI2M2EwYTg

Accessibility

https://accessibility.huit.harvard.edu/digital-accessibility-policy

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story

OPEN ACCESS Freely available online

PLOS | ONE

Biological Time Series Analysis Using a Context Free Language: Applicability to Pulsatile Hormone Data

Dennis A. Dean II[1,2,3,4]*, Gail K. Adler[4,5], David P. Nguyen[2,6], Elizabeth B. Klerman[1,4,5]

1 Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, Massachusetts, United States of America, **2** Neuroscience Statistical Research Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, **3** Biomedical Engineering and Biotechnology Program, University of Massachusetts, Lowell, Massachusetts, United States of America, **4** Harvard Medical School, Boston, Massachusetts, United States of America, **5** Division of Endocrinology, Diabetes and Hypertension, Brigham and Women's Hospital, Boston, Massachusetts, United States of America, **6** National Institute on Aging, National Institutes of Health, Baltimore, Maryland, United States of America

Abstract

We present a novel approach for analyzing biological time-series data using a context-free language (CFL) representation that allows the extraction and quantification of important features from the time-series. This representation results in Hierarchically AdaPtive (HAP) analysis, a suite of multiple complementary techniques that enable rapid analysis of data and does not require the user to set parameters. HAP analysis generates hierarchically organized parameter distributions that allow multi-scale components of the time-series to be quantified and includes a data analysis pipeline that applies recursive analyses to generate hierarchically organized results that extend traditional outcome measures such as pharmacokinetics and inter-pulse interval. Pulsicons, a novel text-based time-series representation also derived from the CFL approach, are introduced as an objective qualitative comparison nomenclature. We apply HAP to the analysis of 24 hours of frequently sampled pulsatile cortisol hormone data, which has known analysis challenges, from 14 healthy women. HAP analysis generated results in seconds and produced dozens of figures for each participant. The results quantify the observed qualitative features of cortisol data as a series of pulse clusters, each consisting of one or more embedded pulses, and identify two ultradian phenotypes in this dataset. HAP analysis is designed to be robust to individual differences and to missing data and may be applied to other pulsatile hormones. Future work can extend HAP analysis to other time-series data types, including oscillatory and other periodic physiological signals.

Citation: Dean DA II, Adler GK, Nguyen DP, Klerman EB (2014) Biological Time Series Analysis Using a Context Free Language: Applicability to Pulsatile Hormone Data. PLoS ONE 9(9): e104087. doi:10.1371/journal.pone.0104087

Editor: Ioannis P. Androulakis, Rutgers University, United States of America

Received September 27, 2012; **Accepted** July 10, 2014; **Published** September 3, 2014

Copyright: © 2014 Dean et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work received funding from the following sources: AFOSR F49620-95-1-0388, FA9550-06-0080; NIH NCRR-GCRC-M01-RR-02635, P01-AG09975, T32 HL07901-10 (DAD), F31-GM095340-01 (DAD), K24-HL105664 (EBK), K02-HD045459 (EBK), RC2-HL101340, R01-AR43130 (GKA), K24-HL103845 (GKA), R01-MH071847, R01 HL098433, and R01 HL098433-02S1(DAD), NSBRI through NASA NCC 9-58 HFP01603 and HPF00405. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: EBK received investigator-initiated research support from Respironics and received an unrestricted gift to the BWH from Sony Corporation. GKA received investigator-initiated support from Novartis. These funding sources do not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials. The other authors have declared that no competing interests exist.

* Email: ddean@research.bwh.harvard.edu

Introduction

Extracting physiologically relevant features and understanding the control underlying biological time-series data is challenging due to potential multiple time-scale components, non-linear relationships among physiological measures, and feed-forward/feed-back control of the system [1]. Analyzing cortisol pulsatility is challenging due to multiple endocrine systems each with its own regulatory mechanisms including circadian modulation and cortisol feedback on the hypothalamus and pituitary gland, [2–4]. Additional analysis challenges include hormone assay error, limitations in hormone sampling rate and individual biological variations. Pulse identification and physiologically based modeling techniques have been developed to quantify cortisol pharmacokinetic parameters and to test plausible physiological mechanisms controlling cortisol pulsatility [5–25]. Limitations to these analysis techniques include the need for the user to choose parameters before beginning analysis, the presence of simplifying assumptions that may not be appropriate for the data set (e.g. homogenous compartments), and the need for different methods for different hormones. Our proposed novel analysis is presented to overcome these limitations.

In addition, our method is designed to select and quantify key qualitative features originally identified during visual inspection of frequently-sampled cortisol hormone time-series data. These qualitative features include differences in pulse frequency and amplitude during sleep and wake, and inter-individual differences in circadian variation in cortisol pulse amplitude. A motivating observation was that these qualitative differences could be explained by the concept of hierarchically organized rises and falls in the data independent of cause (e.g. sleep-wake state, assay error and difference in the signal to secrete). The goal was to develop a technique that required a minimum number of assumptions and that could be applied across a wide number of conditions including different populations including both healthy controls and individuals with altered hormone concentration profiles.

We present our conceptual framework by applying it to cortisol time-series data (**Figure 1**). In this paper, we demonstrate that HAP analysis of cortisol data (a) identifies salient events in multiple

time-scales (e.g., pulses in one time-scale and trends in another), (b) quantifies accumulation and dissipation rates, some of which correspond to standard pharmacokinetic parameters, and (c) characterizes a hierarchical organization of the data. Although HAP is motivated by cortisol concentration profiles, the analysis structure is general enough to apply to other signals.

Methods

Overview

We propose a context free language (CFL) data representation [26–28] for curating and analyzing pulsatile biological and non-biological data. The CFL is a 'model free' approach and is used to identify features in the data series. CFLs are commonly used to specify the structure of programming languages and for natural language processing [29–31] (See **Text S1** Section A for a brief introduction, **Figure S1** for a CFL example, and **Figure S2** for a CFL string processing example). Our novel CFL framework, Hierarchically AdaPtive (HAP) analysis, converts the pulsatile time-series into a symbolic form that allows for quantitative and qualitative analyses based on novel text-based and graph-based representations of the time-series. Collectively, the tools and techniques generate rapid semi-automatic analyses of data.

A CFL-derived algorithm for data processing is defined and allows for analyses of individual and group data. The CFL is designed to capture the multi-scale features present in embedded hormone pulses: one or more secretory episodes are embedded within a longer time-scale and a larger concentration-scale rise and fall in hormone concentration values. Our CFL-based framework is extendable to oscillatory and other periodic biological and non-

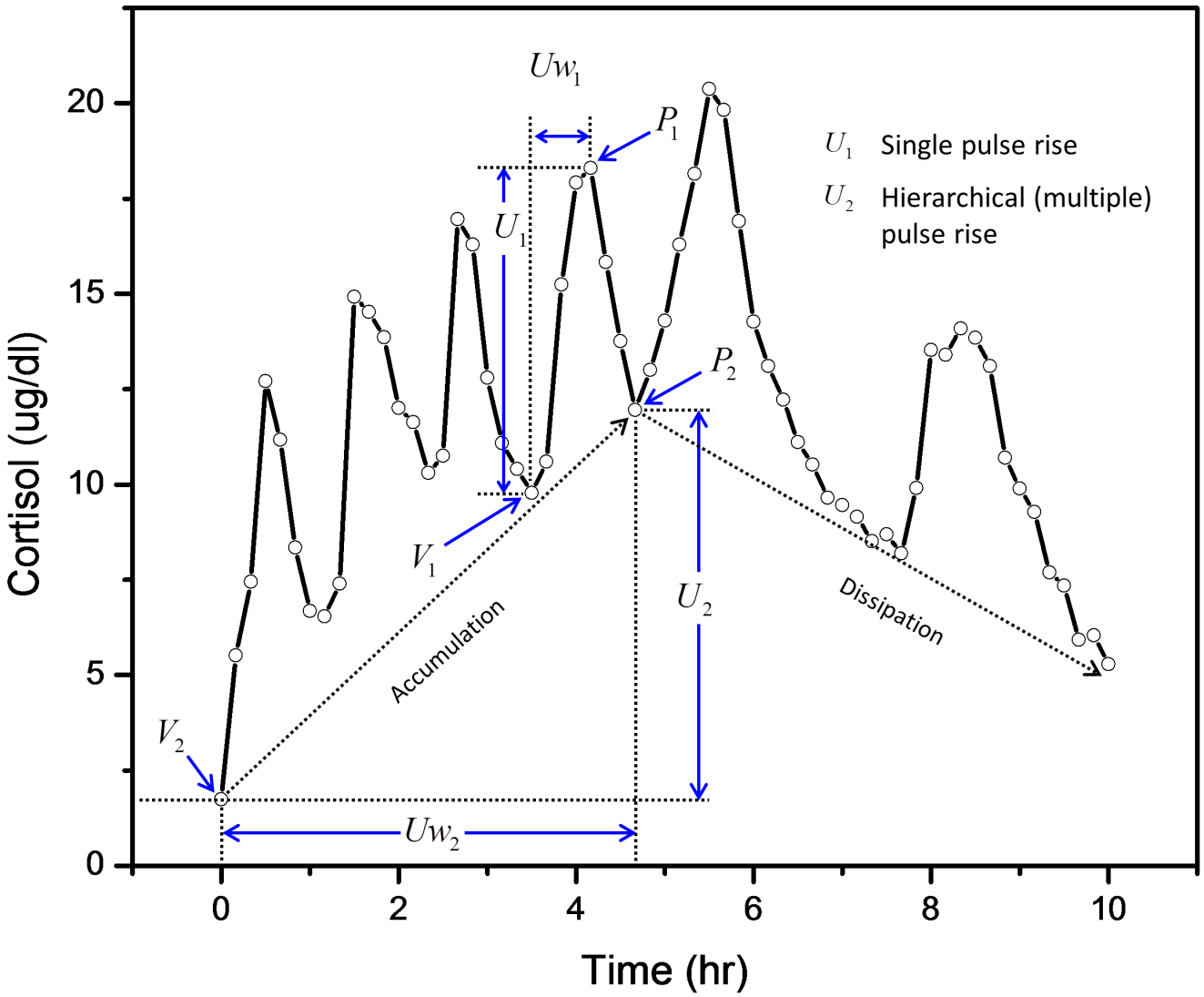

Figure 1. Hierarchically embedded cortisol pulses. An example adapted from experimental data (black line) with six major cortisol pulses. For a **Single Pulse Rise (U_1)** starting at time ~3.5 hours, the single pulse peak is identified with P_1 and the starting nadir/valley is identified with a V_1. The pulse rise is defined as the rise in cortisol concentration from the nadir (V_1) to peak (P_1). The pulse rise time (Uw_1) is the time required for the cortisol concentration levels to rise from a sequence of pulses starting with the local nadir (V_1) to the peak (P_1). The **Hierarchical (multiple) Pulse Rise (U_2)** occurs as the change in concentration from the first nadir (V_2) to last nadir (P_2), which is a local maximum nadir, in the rise portion of this hierarchically organized segment. The hierarchical pulse rise time (Uw_2) is the elapsed time between V_2 and P_2 of the rising portion. The effect of the hierarchical pulses between V_2 and P_2 is an accumulation or increase in cortisol. Similarly, sequences of pulses associated with the dissipation or decrease in cortisol begin at time ~4.8 hours.
doi:10.1371/journal.pone.0104087.g001

biological signals and can be used to quantify and understand underlying mechanisms. The rapid and semi-automatic approach within HAP analysis is also of general scientific importance due to the need for analytic tools that can be applied to the increasing size and number of data sets available.

HAP is presented as suite of integrated tools that characterize individual and group time-series. It contains more than 20,000 lines of MATLAB code that result in an analysis pipeline. One tool converts the data into an appropriate MATLAB structure to allow users to quickly use the other components of HAP. Pulsicons, a signal text representation, provides a framework for pattern analysis; the hierarchical aspect of the pulsicons allows identification of embedded patterns. Lastly, the analysis pipeline is developed with modern software techniques that include versioning and testing of core routines. For example, core routines are written as classes. Using classes is an approach to writing software that allows the core methods to be modified quickly. These software development techniques are recommended by United States Department of Health and Human Services for computer systems used in clinical investigation [32]. Thus, these software tools are structured to enable extensions to include existing and new analysis techniques within the analysis pipeline. HAP is also designed to be complementary to existing data analysis and modeling techniques. For example, its outputs can be used as input to other methods.

Analytical Methods –HAP Analysis

HAP analysis (**Figure 2**) is an analysis framework that includes theoretical and computational steps. The theoretical steps required to generate the computational steps are:

Step 1: A CFL is defined to represent pulsatility within a time-series.

Step 2: A data parser is defined to convert the time-series into a symbolic representation.

Step 3: A production graph is generated from the symbolic representation of data.

Step 4: A text representation of the time-series (pulsicons) is generated from the production graph.

Step 5: Hierarchically organized parameters are computed from traditional equations integrated within the data parser (Domain Specific Step). Note that this is the only domain specific step; it will need to be modified for each type of data.

Details of these methods are described in the **Results** section since they are major results of the paper.

Software

HAP Software. HAP analysis was implemented as a collection of software tools assembled into a computational pipeline. The components of the pipeline include: (1) a general hormone database structure for organizing hormone time-series data and experiment details including experimental protocol, participant demographics and sleep-wake state, (2) tools for applying HAP to data stored in the database, (3) tools for visualizing the hormone data contained in the database, (4) tools for visualizing the results from the HAP analysis, and (5) tools for creating summaries of individual and group data results.

The HAP analysis pipeline was written with MATLAB ver. 7.3 and contains 87 MATLAB script files. The Graphical User's Interface Development Environment (Guide) toolbox was used to create an interface for selecting, visualizing, and analyzing the hormone database structure. The HAP_Analysis program used in this analysis and the source code can be found at https://github.com/DennisDean/HapSource/releases. HAP_Analysis is a windows executable and does not require MATLAB to be installed on the computer. **Text S1** contains additional program details and a summary of figures that can be generated by the program is shown in **Figure S3**. A video of the program executing an analysis is posted on YouTube (http://youtu.be/ggJDK7alI2M).

Cortisol Model Simulation Software. A published mathematical model of cortisol concentration levels [33] was implemented and used to test the HAP implementation. Features of the mathematical model include randomly sampled interpulse intervals with circadian amplitude modulation. The MATLAB simulation script and a test program can be found at https://github.com/DennisDean/BrownCortisolModel. A brief description of the model and software is included in **Text S1**.

Data

Experimental Data. Fourteen healthy women were studied in an inpatient protocol [34]. Blood was sampled through an indwelling intravenous catheter every ten minutes for 24 hours and later assayed for cortisol. The first eight hours consisted of a scheduled sleep period and the remaining 16-hour wake period employed a constant routine protocol during which participants remained awake in a semi-recumbent position in dim lights and were given small regularly spaced meals [35–37].

Ethics Statement

All participants gave informed consent, and the experimental protocol was approved by the Partners Healthcare Institutional Review Board.

Results

The derivation for steps 1–4 of HAP analysis are presented first, followed by their application to cortisol data.

Theoretical Results

Context Free Language for Identifying Hierarchically Organized Pulses (Step 1). The CFL was motivated by the observation that a rise or fall in the cortisol concentration profile can be due to an embedded set of pulses. This multi-scale embedded feature can be represented as a hierarchical structure. The detailed construction and steps involved in a CFL example of embedded parentheses is included in Section A of **Text S1**. The details of the CFL for HAP are presented here.

Let L_C be the set of the CFL that can identify hierarchical rises and falls in a time-series representing cortisol concentration measurements. L_C is composed of a grammar that can generate strings to represent any valid cortisol concentration or other pulsatile time-series; these strings can be transformed into a computation for recognizing the CFL directly from the pulsatile time-series.

The CFL grammar that generates these strings in the language L_C is composed of a quadruple, where a quadruple is a set of four definitions, that includes the alphabet $V \rightarrow \{(,)\}$, the set of terminals $\Sigma \rightarrow \{(,)\}$, the set of rules $R \rightarrow \{S \rightarrow SS, S \rightarrow (S), S \rightarrow (\ \ ), S \rightarrow S : S\}$, and the start symbol $S \rightarrow \{S\}$. The components of the CFL grammar are shown in Equations (1)–(4).

$$V \rightarrow \{(\ : \ )\} \tag{1}$$

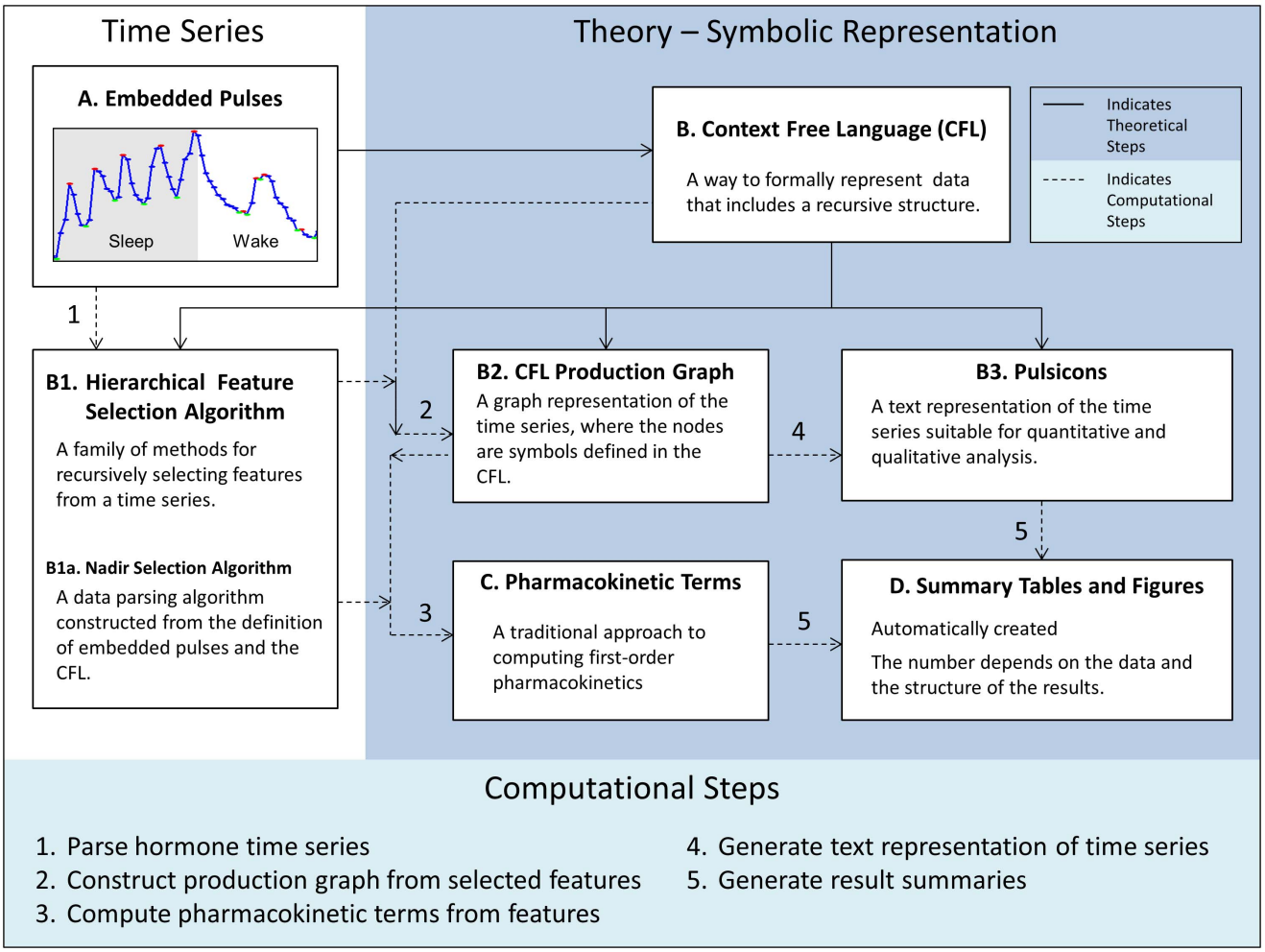

Figure 2. Hierarchically AdaPtive (HAP) Analysis Schematic. The figure is divided into 3 sections: Time-series (left panel), Theory – Symbolic Representation (right panel) and Computational Steps (bottom panel). Solid arrows indicate theoretical steps; dotted lines with numbers indicate computational steps listed in the bottom panel.
doi:10.1371/journal.pone.0104087.g002

$$\Sigma \rightarrow \{(\ : \ )\} \tag{2}$$

$$R \rightarrow \left\{ \begin{array}{lcl} S & \rightarrow & SS \\ S & \rightarrow & (S) \\ S & \rightarrow & () \\ S & \rightarrow & S:S \end{array} \right\} \tag{3}$$

$$S \rightarrow \{S\} \tag{4}$$

This CFL provides a symbolic and compact representation of hierarchically organized pulses. Thus, the CFL is a way to represent pulsatile time-series by discrete sections or tokens, where each section describes features present in the data as defined in the language's alphabet.

Nadir Selection Algorithm Specification (Step 2). *CFL and Nadir Selection Algorithm Relationship*. The CFL described above is a compact representation of pulsatile data and can be used with time-series. However, direct application is problematic since it is not possible to deterministically identify the next valid CFL production when examining a portion of the time-series from left to right (i.e., earlier to later in time). Additional information is required to determine the next valid production according to the theory of computation [26,28]. The nondeterministic nature or use of a push-down automata (i.e., embedded parentheses example in Section A of **Text S1**), which is the usual means of analyzing a CFL, may require more computational resources than a deterministic algorithm, and therefore computation speed can be slow. Instead the inherent structure in matched parentheses can be used to construct an efficient and deterministic algorithm for recognizing the CFL. For these pulsatile data, a deterministic partitioning algorithm is used. The idea behind our specific partitioning algorithm, the Nadir Selection Algorithm, is that processing the entire time-series simultaneously allows matching pairs of rises, peaks, and falls in the data to be determined recursively. Thus, the Nadir Selection Algorithm is a time-series parser for converting the data into a symbolic representation. In addition, since we observed that the relationship between time and amplitude of nadirs in 24 hour cortisol profiles varied substantially by individuals (details below), we conjectured that this context free representation of pulsatility could be useful as a non-parametric

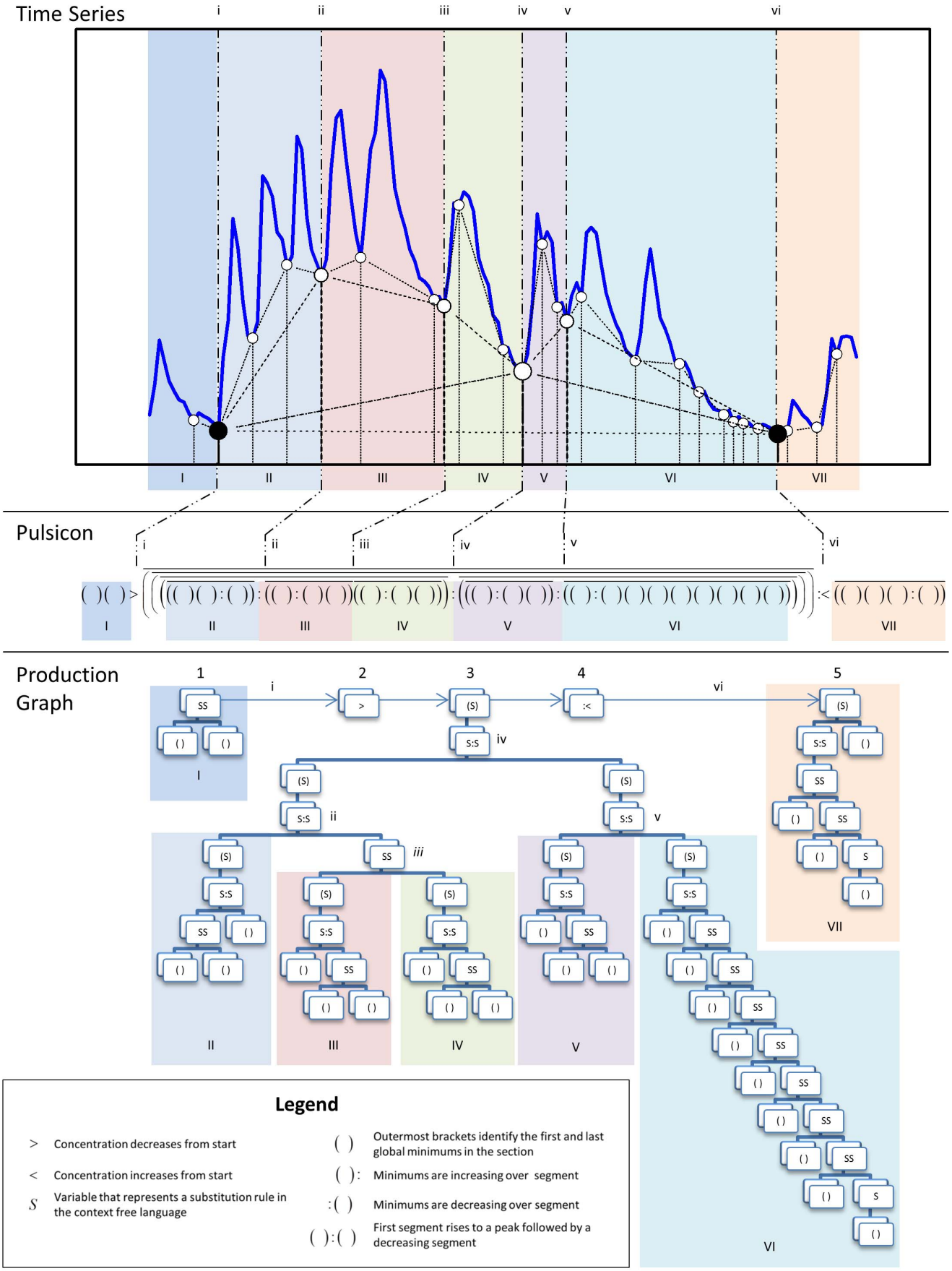
Time Series
i
ii
iii
iv
v
vi
I
II
III
IV
V
VI
VII
Pulsicon
Production Graph
1
2
3
4
5
SS
S:S
(S)
S
Legend
> Concentration decreases from start
< Concentration increases from start
S Variable that represents a substitution rule in the context free language
( ) Outermost brackets identify the first and last global minimums in the section
( ): Minimums are increasing over segment
:( ) Minimums are decreasing over segment
( ):( ) First segment rises to a peak followed by a decreasing segment

Figure 3. Qualitative time-series representations for a single participant (ID6). The dotted vertical lines with lower case Roman numerals represent the same time points in the Time-Series, the Pulsicon, and the Production Graph for participant ID6. Shaded rectangles labeled with uppercase roman numerals represent the same temporal interval in the Time-Series, the Pulsicon, and the Production Graph. **Time-Series:** The 24-hour cortisol time-series. Nadirs identified in each of the four HAP iterations are connected with lines. The terminal HAP nadirs are identified with black circles. **Pulsicon:** The pulsicon for the time-series. **Production Graph:** Each rectangle represents productions: the application of rules defined in the CFL The five rectangular boxes represent the key features of the time-series. The first two rectangles (Labeled 1 and 2) show that there is a decreasing portion prior to the main portion of the cortisol time-series with the "$>$" in rectangle 2 (e.g., before left black circle). The main portion of the cortisol time-series is represented by the center rectangle (Labeled 3) with the production symbol (S). The production symbol (S) is a variable that can be replaced with any of the rules defined in the context free language. This section contains a hierarchy of embedded pulses. The last two rectangles (Labeled 4 and 5) from the production graph show that the cortisol time-series is rising after the main portion (rectangle labeled 3) of the cortisol time-series (e.g., after right black circle).
doi:10.1371/journal.pone.0104087.g003

approach to represent the multi-scale relationship of features (e.g., nadirs and peaks) in a pulsatile time-series. Note that an equally valid algorithm anchored on peaks could have been derived. The nadirs in the time-series were selected to anchor the partitioning algorithm since hormone pulse start times are traditionally associated with nadirs.

Nadir Selection Algorithm. The Nadir Selection Algorithm is based on a recursive algorithm for identifying and classifying peaks and troughs within the time-series data. The first step of the algorithm is to identify local peaks and nadirs. The algorithm is recursively called with nadirs as input until no rises or falls in the data are identified (**Figure 3**). The recursion results are used to reconstruct the productions (i.e., rises and falls) required to generate the time-series. Quantitative comparisons identify relative changes within the time-series within this hierarchy of recursion results. The quantitative comparisons include expressions for identifying concentration rises and falls, plateaus, flat regions, and missing data. The recursive algorithm was designed to exploit parallelism in the data and results in an efficient method for processing data [27,38–40].

Let $\vec{Y}=\left[\vec{T},\vec{C}\right]$, where the time-series $\vec{Y}$ is composed of the time vector $\vec{T}$ and the concentration vector $\vec{C}$. The time vector $\vec{T}$ is the set of concentration sampling times $\{t_1,t_2,\ldots,t_N\}$, assumed to be sampled at a constant interval (Δt), and may have missing data points. Time-based equations are presented in equations (5) and (6).

$$\vec{T}=\{t_1,t_2,\ldots,t_N\} \tag{5}$$

$$\forall(t_{i+1}-t_i)\rightarrow L\cdot\Delta t \text{ where } L\in\mathbb{Z} \tag{6}$$

Let $\vec{C}$ be the set of hormone concentrations values $\{c_1,c_2,\ldots c_N\}$ that correspond to the sampling times $\vec{t}$ as shown in equation (7).

$$C=\{c_1,c_2,\ldots,c_N\} \tag{7}$$

A comparator operator (J) which is a function of the concentration data (C) and a set of comparators (O) is defined in equations (8) and (9). The comparator is a tuple where each entry is a member of the set of simple comparators defined as $\{<,>,=\}$. Specific pairs of operators are selected that correspond to visually relevant features of hormone concentration data relative to the center point, namely nadirs ($<,<$), peaks ($>,>$), rising plateaus ($>,=$), descending plateaus ($<,=$), and flat regions ($=,=$). Details regarding the interpretation of the comparator operators are described in Section B of **Text S1** and the comparator operators are summarized in **Figure S4**.

$$\begin{aligned}&\vec{J}(C)\rightarrow\vec{J}(c_i,\vec{o}_a)\rightarrow\\&\left[\left(c_{i+1}o_{a_1}c_i\right),\left(c_{i+1}o_{a_2}c_{i+2}\right)\right]\forall i,1\le i\le N;\forall a,o_a\in O\end{aligned} \tag{8}$$

$$\begin{aligned}&\vec{O}\rightarrow\{\vec{o}_1,\vec{o}_2,\ldots,\vec{o}_9\}\rightarrow\\&\left\{\begin{matrix}(<,<)(>,>)(=,=)(<,>)(>,<)\\(=,>)(=,<)(<,=)(>,=)\end{matrix}\right\}\end{aligned} \tag{9}$$

Let $\vec{A}$ correspond to the features that are identifiable in the data as shown in equation (10):

$$A\rightarrow\{a_1,a_2,\ldots,a_9\}\rightarrow\{P,N,F,R,D,DF,RF,FR,FD\} \tag{10}$$

where each feature a_1 corresponds to the following one of the 9 operators described above.

Let $\vec{A}_0$ correspond to the features identified in the data by applying the comparator operator (J) with the set of operators (O) to the concentration data (C).

Let $\vec{T}_0$ and $\vec{C}_O$ correspond respectively to the times and concentration values of the q features identified in $\vec{A}_0$. The features that identified corresponding times and corresponding concentration values are defined by equations (11)–(13).

$$\begin{aligned}&\vec{A}_0\rightarrow\left\{a_{0_1},a_{0_2},\ldots,a_{0_q}\right\}\forall i,a_{0_i}\in J(C,O);\\&\forall i,a_{0_i}\in(C,O);\forall i,a_{0_i}\in\vec{A}\end{aligned} \tag{11}$$

$$\vec{T}_O\rightarrow\{t_1,t_2,\ldots,t_q\}\forall i,t_i\in T;t_i<t_{i+1} \tag{12}$$

$$\vec{C}_0\rightarrow\{c_1,c_2,\ldots,c_q\}\forall i,c_i\in C \tag{13}$$

A family of recursions is defined by recursing on a subset of the identified features. A subscripted S is used to designate the selected features with corresponding times and corresponding concentration values as shown in equations (14)–(16).

$$\vec{A}_{0_S}\rightarrow\left\{a_{0_{s1}},a_{0_{s2}},\ldots,a_{0_{sr}}\right\}\forall i,a_{0_{si}}\in\vec{A} \tag{14}$$

$$\vec{T}_{0_S}\rightarrow\{t_1,t_2,\ldots,t_q\}\forall i,t_i\in;t_i<t_{i+1} \tag{15}$$

$$\vec{C}_{0_S} \rightarrow \{c_1, c_2, \ldots, c_q\} \forall i, c_i \in C \tag{16}$$

The subset $\vec{J}(C_{O_S})$ is defined to be the nadirs for the purpose of partitioning the hormone time-series into distinct pulsatile regions as described in equation (17).

$$\vec{J}\left(C_{0_S}\right) \rightarrow \vec{J}(c_i, \vec{o}_a) \rightarrow \left[\left(c_{i+1} o_{a_1} c_i\right), \left(c_{i+1} o_{a_2} c_{1+2}\right)\right] \forall i,$$
$$1 \leq i \leq N; \forall a, o_a \in O \tag{17}$$

Similarly, a partitioning from peak to peak can be defined by selecting the peak as the subset.

Algorithm Interpretation. The identified nadirs have both a mathematical and physiological interpretation. At each level of iteration, two adjacent vertical lines in **Figure 3** identify a nadir-nadir interval with a single intervening peak; these HAP-selected nadirs identify embedded rising and falling concentration segments. Mathematically, the same concept is referred to as the forced state of the system [41]. Nadirs selected at multiple iterations are identified as fixed points in the sequential dynamical systems literature [42–44]. The selected nadirs correspond to the start of an increase in cortisol concentration and are likely to correspond to the approximate timing of a signal to secrete cortisol.

Computing Accumulation Rates, Dissipation Rates, and Inter-nadir Intervals. The extracted hierarchical features provide a concise definition of pulsatility that can be used to calculate traditional hormone pharmacokinetic measures. In this section, we present definitions of accumulation rates, dissipation rates and inter-nadir intervals derived directly from features extracted from HAP. These metrics are analogous to secretory rates, clearance rates, and interpulse intervals, respectively,

The change in cortisol concentration rate (ΔC) over a finite time interval $\mathrm{V}t_C$ is equal to the starting concentration level (C_1) plus the cortisol secreted during the interval minus the cortisol eliminated during the interval. The cortisol generated during the interval is equal to the secretory rate (β_I) during secretion multiplied by the change in time (Δt_S). Similarly, the cortisol dissipated is equal to the clearance rate (β_C) multiplied by the time cortisol is being cleared from the change in time Δt_C. This relationship is shown in equation (18).

$$\Delta C = C_2 - C_1 = \beta(\Delta t_S) - \beta(\Delta t_C) \tag{18}$$

For a simple pulse, the duration of secretion is equivalent to the rise duration while duration of clearance is equivalent to the fall duration, as in equation (19).

$$\begin{aligned} \Delta t_S &= t_{S_b} - t_{S_a} &&= \text{rise duration} \\ \Delta t_C &= t_{C_b} - t_{C_a} &&= \text{fall duration} \end{aligned} \tag{19}$$

t_{S_b} and t_{S_a} are the end and start of the rise interval, respectively. Similarly, t_{C_b} and t_{C_a} are the end and start of the fall interval, respectively. For simplicity, secretion and clearance are assumed to be temporally distinct.

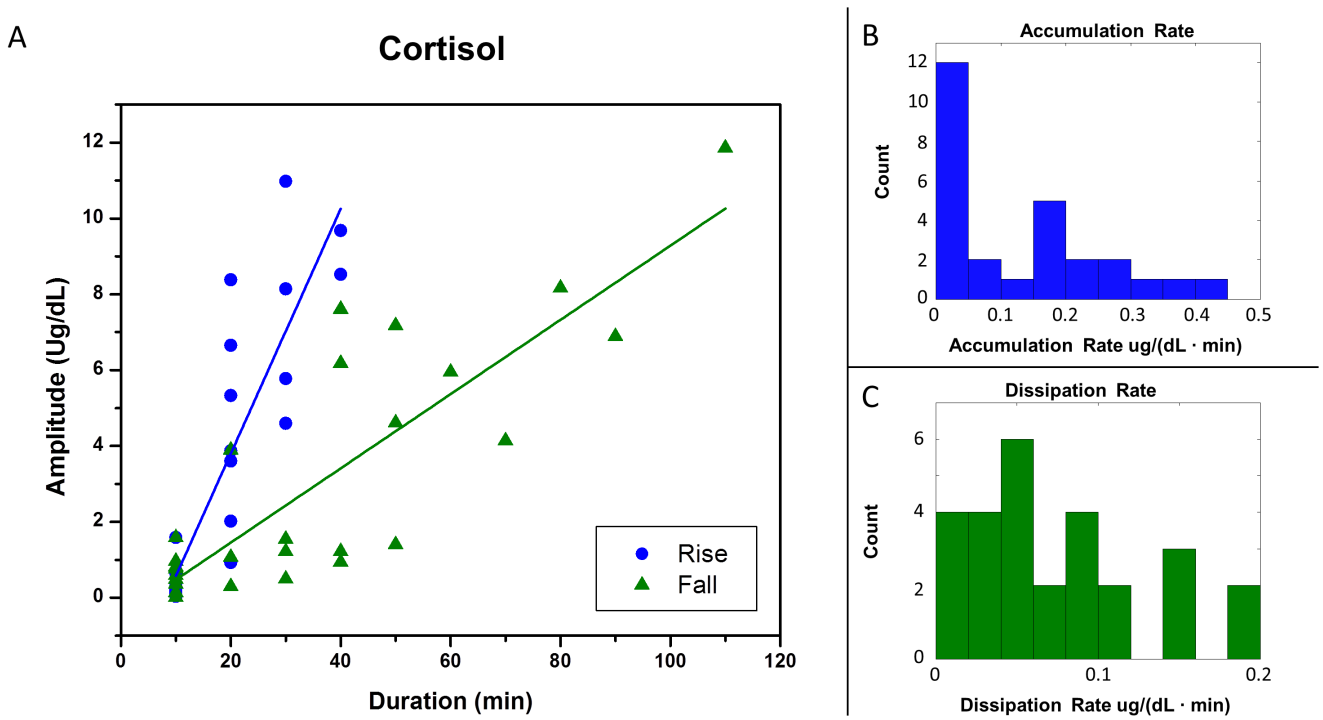

Figure 4. Accumulation and dissipation rates for a single participant (ID6). (**A**) Amplitude as a function of the rise or fall duration extracted from the first iteration of the HAP algorithm for participant ID6. Concentration rises (changes between local nadirs and subsequent local peaks) are shown with blue circles. Concentration falls (changes between local peaks and subsequent local nadirs) are shown with green triangles. Regression lines are shown separately for the rises (blue circles) and falls (green triangles). (**B**) A histogram of the instantaneous accumulation rates (=rise amplitude/rise duration) from Panel A. (**C**) A histogram of the instantaneous dissipation rates (=fall amplitude/fall duration) from Panel A.
doi:10.1371/journal.pone.0104087.g004

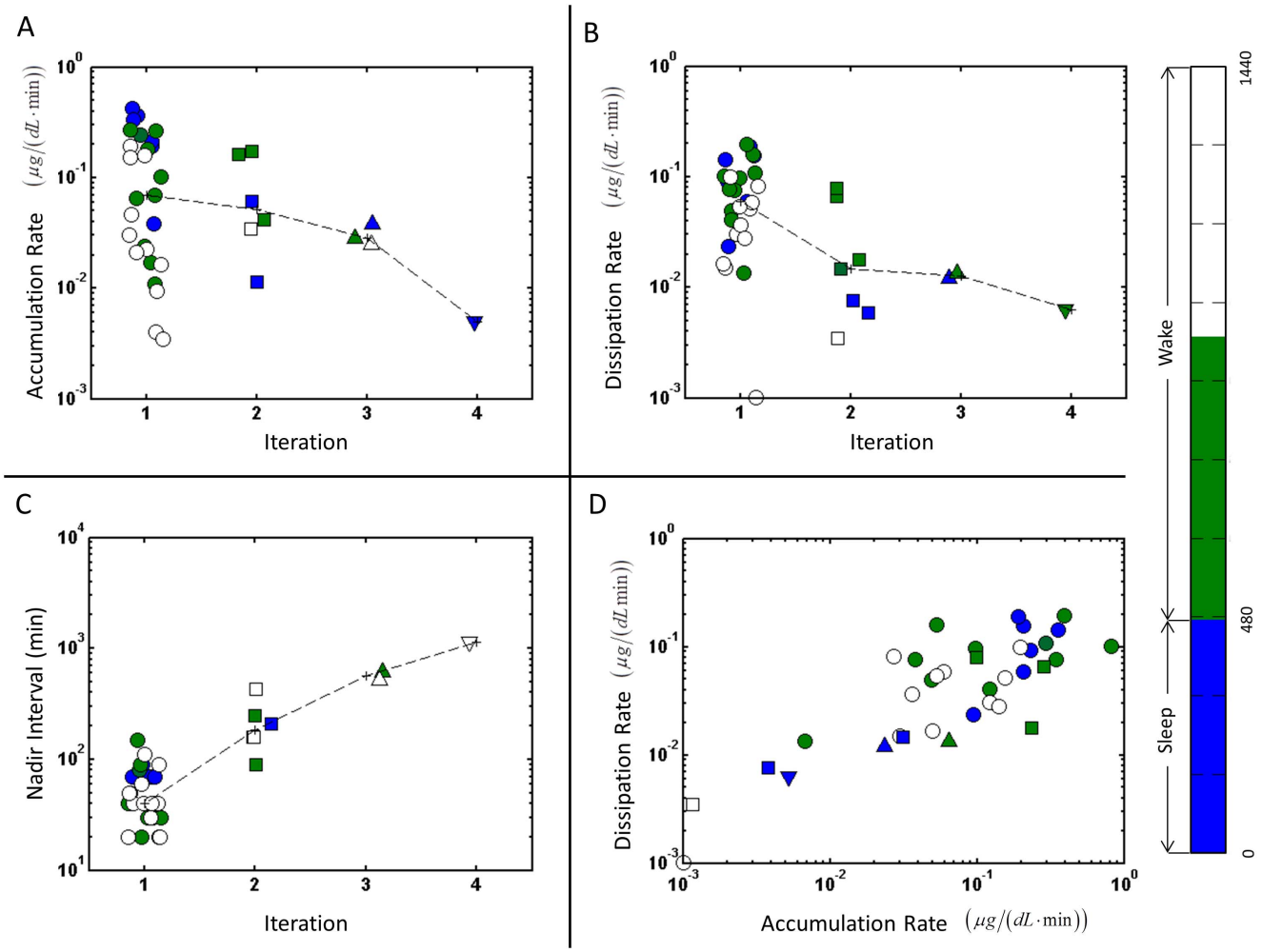

Figure 5. Multi-scale analysis of features identified with HAP from a single participant (ID6). Each analysis iteration for participant ID6 is shown with a different symbol: 1-circle, 2-square, 3-up triangle, 4-down triangle. The symbol color indicates the timing of the event with the color code in right-most vertical panel: the sleep episode (blue), the first 8 hours of wake (green), or second 8 hours of wake (white). The dashed lines connect median value between iterations. (**A**) Cortisol accumulation rates for nadirs identified at each iteration. Accumulation rates for the first iteration are analogous to secretory rates. (**B**) Cortisol dissipation rates for nadirs identified at each iteration. Dissipation rates for the first iteration are analogous to clearance rates. (**C**) The nadir intervals identified at each iteration are shown. The event intervals at the first iteration are analogous to the inter-pulse interval. (**D**) The relationship between an accumulation rate and the immediately following dissipation rate.
doi:10.1371/journal.pone.0104087.g005

Consider a subset of features from $\vec{A}_i$ selected at iteration i where every entry is a nadir (N), decreasing to flat (DF), or flat to rising (FR). Then there exists a single peak (P), rising to flat (RF), or flat to decreasing (FD) within $\vec{A}_i$ that provides sufficient information to estimate the first order secretory and clearance rates. For simplicity, consider a subset of features that contains three entries sequentially consisting of a nadir, peak, and nadir as shown in equation (20).

$$\vec{A}_{i_S} = \left\{a_{i_1}, a_{i_2}, a_{i_3}\right\}, a_{i_1} \in \{N, DF, FR\}, \quad a_{i_2} \in \{P, RF, FD\}, a_{i_3} \in \{N, DF, FR\} \tag{20}$$

If we assume that the effective secretory period corresponds to the rising portion and the clearance period corresponds to the descending portion of the interval, then equations (11)–(13), (18), and (19) can be used to compute the first order secretory rate S_{i_k}, infusion rate C_{i_k}, and inter-nadir interval IP_{i_k} directly from HAP outputs as shown in equations (21)–(23).

$$(\beta_I)_{i_k} = \left(C_{i_{k+1}} - C_{i_k}\right) \Big/ \left(t_{i_{k+1}} - t_{i_k}\right) \tag{21}$$

$$(\beta_C)_{i_k} = \left(C_{i_{k+1}} - C_{i_{k+2}}\right) \Big/ \left(t_{i_{k+2}} - t_{i_{k+1}}\right) \tag{22}$$

$$IP_{i_k} = t_{i_{k+2}} - t_{i_k} \tag{23}$$

For the simple example above, there is only one secretory rate, one clearance rate, and one interpulse interval because there are no embedded pulses. For a subset of features $\vec{A}_{i_S} = \{a_{i_1}, a_{i_2}, \ldots, a_{i_M}\}$ where each entry ($\forall a_i \in \{N, DF, FR\}$), then there will be k secretory rates, k clearance rates, and k interpulse

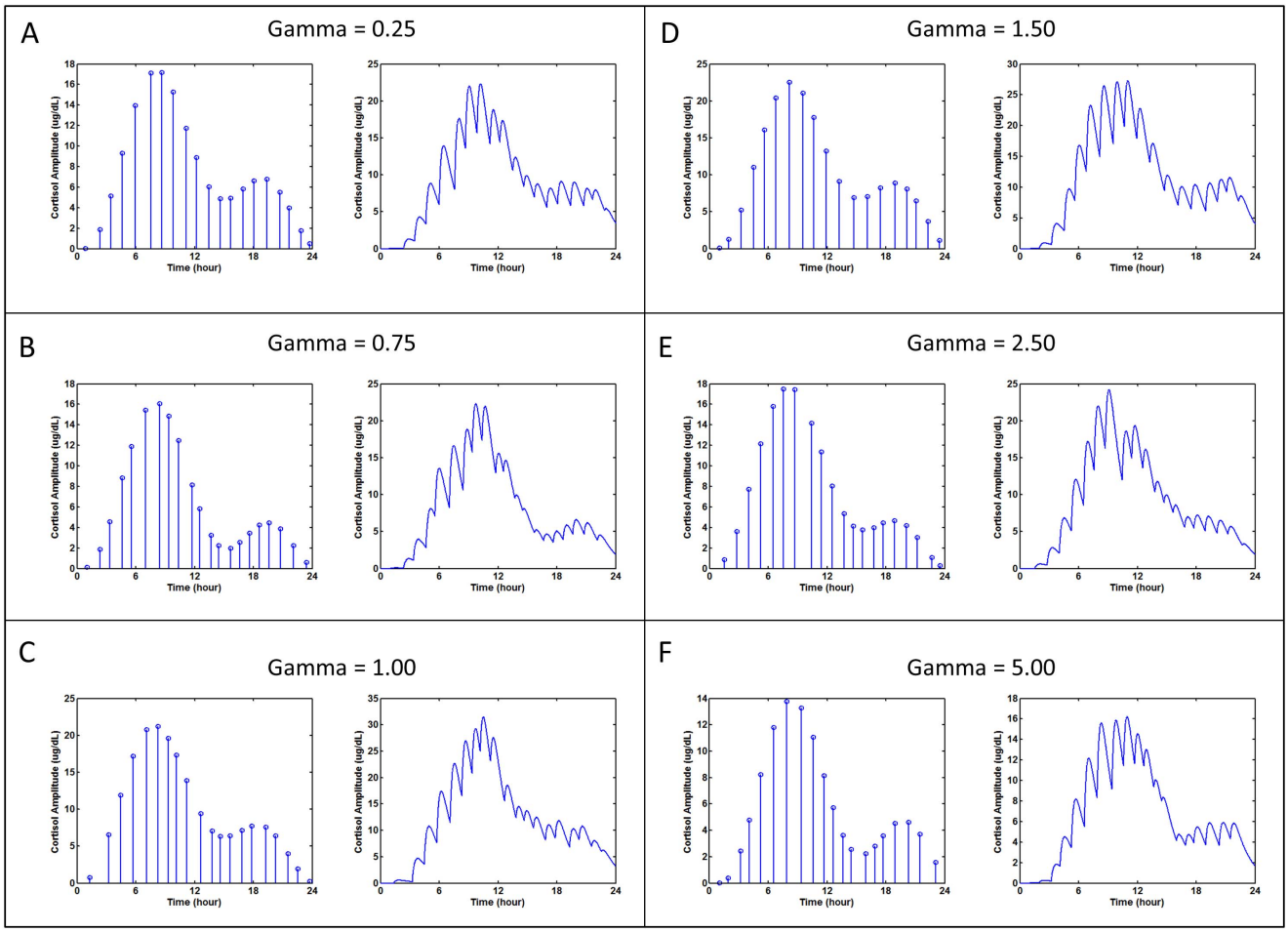

Figure 6. Simulated 24-hour cortisol concentration time-series profiles with gamma values from Table 1. Within each panel, the left figure plots the time and amplitudes of each pulse and the right figure plots the resulting simulated dataset. The gamma value affects the clearance rate. (A) Gamma = 0.25 (B) Gamma = 0.75 (C) Gamma = 1.0 (D) Gamma = 1.5 (E) Gamma = 2.5, (F) Gamma = 5.0.
doi:10.1371/journal.pone.0104087.g006

intervals where k is equal to the number of local nadirs minus one. We define the hierarchical secretion rates as accumulation rates, the hierarchical clearance rates as dissipation rates and hierarchical interpulse intervals as inter-nadir intervals.

Qualitative Time-Series Representation

The HAP analysis framework includes two visual time-series representations: production graphs (Step 3) and a text representation we term pulsicons (Step 4).

Qualitative Features Present in the Production Graph (Step 3). The CFL production graph is presented as a way to describe the qualitative features present in a cortisol time-series. The CFL production graph is a summary of the steps in the language required to represent the data. The goal of the graph is to better visualize qualitative features such as the relations of embedded pulses and could be used to integrate graph-based analysis algorithms in future analytical work.

The production graph for the example in **Figure 3** is organized in a sequence of 5 rectangles that form a high level description of the time-series from left to right. Three rectangles - labeled 1, 3, and 5 in the production graph - contain the productions required to generate their corresponding pulsicon sections. These three rectangles are expanded to show the productions required to produce the pulsicon section that represents this cortisol time-series. Two rectangles (labeled 2 and 4) identify whether the data are rising or falling relative to the main section (labeled 3). Both points in time and temporal intervals are identified to reinforce the relationship between the time-series, the pulsicon, and the production graph.

Text Representation of a Single Time-Series (Step 4). HAP analysis generates a pulsicon or text description of the time-series. A pulsicon is a complete record of rises and falls within the time-series and can be used to evaluate the relationship among embedded pulses. Pulsicons are composed of a set of characters $\{(,), :, <, >\}$. A "(" indicates the start of rise and a ")" indicates the end of a decreasing pulse. The peak between two nadirs is identified with a ":". The ">" symbol indicates a decreasing section of the time-series and the "<" symbol indicates an increasing section of the time-series. Both ">" and "<" describe data points at either the start or the end of the time-series. A pulsicon is generated for each step in the algorithm and at each hierarchical iteration. The pulsicon generated from the production graph described above is shown in **Figure 3**.

The pulsicon text representation of a single pulsatile time-series is composed of the same alphabet used in the CFL ($V \rightarrow \{(, :,)\}$) and is generated from the production graphs using standard production techniques for generating text strings from a graph representation [26–28]. The pulsicon provides a structure for

Table 1. Amplitude (A) and secretion times (T) for simulated dataset*.

	Gamma = 0.25		Gamma = 0.75		Gamma = 1.0		Gamma = 1.5		Gamma = 2.5		Gamma = 5.0	
Pulse	**A**	**T**	**A**	**T**	**A**	**T**	**A**	**T**	**A**	**T**	**A**	**T**
ID	**μg/dL**	**min**	**μg/dL**	**min**	**μg/dL**	**min**	**μg/dL**	**min**	**μg/dL**	**min**	**μg/dL**	**min**
1	***0.02***	***51***	0.11	60	0.79	78	0.08	62	0.87	90	***0.00***	***60***
2	1.89	140	1.87	140	6.54	194	1.28	117	3.63	166	0.38	114
3	5.13	208	4.54	201	11.89	267	5.21	193	7.75	241	2.45	191
4	9.30	277	8.85	278	17.20	344	11.07	271	12.14	314	4.77	245
5	13.98	358	11.92	333	20.75	426	16.12	336	15.82	389	8.23	314
6	17.15	453	15.43	421	21.22	497	20.39	406	17.52	454	11.82	394
7	17.16	519	16.07	505	19.64	560	22.51	489	17.46	522	13.76	472
8	15.29	590	14.88	561	17.37	610	21.04	571	14.15	626	13.27	563
9	11.73	670	12.47	621	13.89	673	17.77	639	11.35	684	11.06	634
10	8.90	730	8.17	705	9.44	759	13.24	715	8.08	752	8.14	702
11	6.03	809	5.86	752	7.10	829	9.16	795	5.34	826	5.73	758
12	4.88	877	3.24	822	6.38	879	6.90	884	4.16	882	3.62	819
13	4.92	941	2.25	869	6.38	936	7.06	968	3.75	937	2.57	867
14	5.83	1017	1.96	941	7.15	1009	8.24	1048	4.01	1008	2.22	958
15	6.65	1083	2.53	998	7.76	1069	8.94	1137	4.44	1065	2.81	1013
16	6.79	1163	3.44	1057	7.58	1157	8.10	1209	4.68	1133	3.59	1064
17	5.53	1243	4.23	1117	6.45	1215	6.49	1265	4.20	1206	4.52	1137
18	4.00	1298	4.48	1176	3.99	1292	3.67	1338	3.03	1273	4.62	1220
19	1.76	1371	3.88	1246	1.92	1352	1.14	1409	1.10	1361	3.72	1288
20	0.51	1424	2.26	1327	***0.22***	***1424***			***0.27***	***1412***	1.58	1385
21			0.60	1406								

Bolded entries were not identified by HAP Analysis.
doi:10.1371/journal.pone.0104087.t001

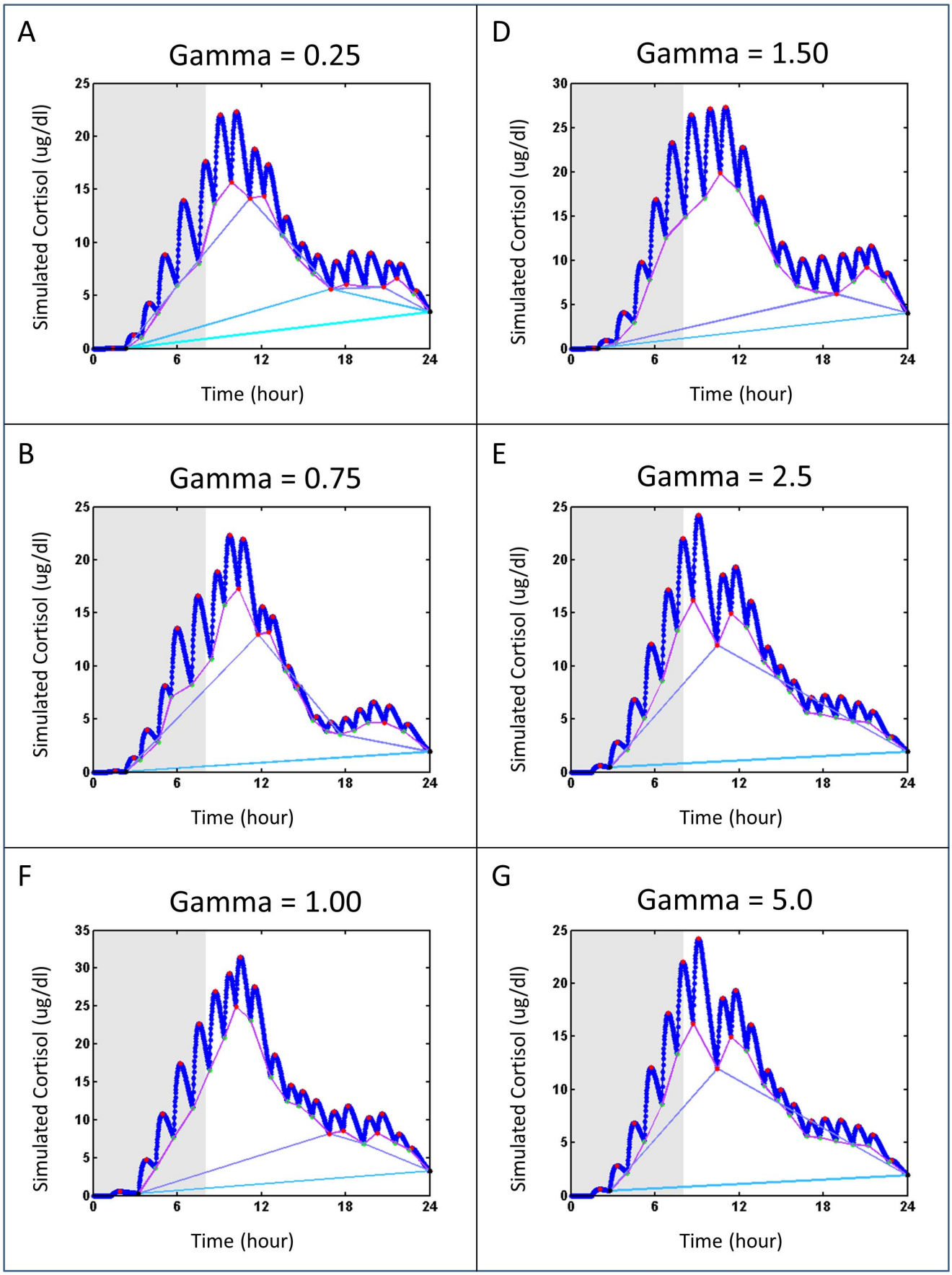
A
Gamma = 0.25
Simulated Cortisol (ug/dl)
Time (hour)
D
Gamma = 1.50
Simulated Cortisol (ug/dl)
Time (hour)
B
Gamma = 0.75
Simulated Cortisol (ug/dl)
Time (hour)
E
Gamma = 2.5
Simulated Cortisol (ug/dl)
Time (hour)
F
Gamma = 1.00
Simulated Cortisol (ug/dl)
Time (hour)
G
Gamma = 5.0
Simulated Cortisol (ug/dl)
Time (hour)

Figure 7. HAP analysis of simulated cortisol test set. (A–F) Simulated data generated from the data sets shown in Fig. 4 panels A–F. Simulated data are identified with blue markers. Peaks identified during the nadir selection algorithm are in red. Nadirs identified at a given recursive step are joined with a line, where line color represents a different recursive step.
doi:10.1371/journal.pone.0104087.g007

summarizing the quantitative attributes of the time-series in a way that can be used by automated processing routines. For example, the pulsicon provides a common interface for automatically generating summary figures about the specific data set (e.g., **Figures 4–5**). Section D of **Text S1** contains (i) an additional pulsicon example and (ii) an example that demonstrates the relationship between production graphs and pulsicons (**Figure S5**).

Application of HAP analysis to simulated data

HAP Analysis output was tested with a simulated data set for which the number, timing, and amplitude of pulses were known. The simulated cortisol test set shown in **Figure 6**. The interpulse interval and circadian amplitude variation were sampled randomly from published distributions with different gamma parameter values. The amplitudes and simulated secretion times are shown in **Table 1**. The gamma parameter is defined as a renewal process that affects the clearance of the simulated pulse. The MATLAB script files used to generate the simulated data can be found at https://github.com/DennisDean/BrownCortisolModel. A brief description the model and location for the simulation code are in Section B of **Text S1**. We used these simulated parameters to validate both the HAP analysis approach and the program implementation

HAP_Analysis Verification. HAP analysis converged for each simulated data set. The program identified rise time and inter-pulse intervals for 97% percent of the pulses (116 of 120, **Table 1, Figure 7**). The pulses missed by the algorithm had secretion amplitudes of less than 0.27 ug/dL. The amplitude of the missed pulses was not sufficient to impact the simulated time series profile. We would not expect HAP to identify these low amplitude events.

Application of HAP analysis to experimental cortisol data (Step 5)

Within seconds HAP generated individualized accumulation and dissipation rates for all participants. **Tables 2** and **3** show results for all participants; single participant results are shown in **Table 4** and **Figures 4–5**. The Nadir Selection Algorithm parsed the data for all participants in either 3 or 4 recursive steps (**Table 2**). In addition, the software pipeline automatically created graphs, summary tables, and text descriptions of each cortisol profile in minutes.

There is a difference in patterns for participant patterns with 3 vs. 4 iterations. Participants with 4 iterations have more pulses that are embedded (i.e., (( ))) while those with 3 iterations have more simple rises and falls (i.e., ( ) ( ) ( )).

As an example, for participant ID6, the Nadir Selection Algorithm identifies 27 peaks and 28 nadirs from the 145 points present in the data at the first recursive step. In the second recursive step, 7 peaks and 7 nadirs were identified. In the third step of the algorithm, 3 peaks and 3 nadirs were identified. In the fourth and terminal recursive step, 1 peak and 1 nadir were identified. (**Table 4**)

Accumulation and Dissipation Rates. The accumulation rate and dissipation rates are calculated from the first HAP iteration (**Figure 4**) for participant ID6. The median accumulation and dissipation rates across all points are 0.067 and 0.059 μg/(dL min), respectively for this individual. Fitting a regression line to the rise points results in a 0.322 μg/(dL min) accumulation rate

Table 2. Pulsicons for All Participants.

Pulsicon by Iteration Level				
ID	**1**	**2**	**3**	**4**
1	(((000:0))((0:00):((0:0)((0:00000))((0:00)((0:0))(0:00000)))	((00:00000))	(())	
2	(((000:0))(000:00):((0:0000)(00:0000000))(0:0)))	((00:000))	(())	
4	(((0:0))((00:00):((00000:00)(0:00)(0:0)((0:000)((0:000)))	((00:00000))	(())	
5	(((000000:0)):((00:0000000)(0000:0000)(0:0000)))	((0:000))	(())	
6	(((000000:0))((0:000)):((0000:000)((0:000000)))	((00:00))	(())	
8	(((00:0))((00:0):((0:00000)(0:00)(0:00000000))(0:0)))	((00:0000))	(())	
11	(((0:0))((00:00):((0:0000)((0:000)((00:00000000)))	((00:000))	(())	
12	(((0:0))((000:00):((0:0000)((00:000)((0:00)))	((00:000))	(())	
13	(((00:0))((00:0))(0:00):((0:0)((0:00)((0:000)))	((000:000))	(())	
3	((((00:0))(00:00)(0:00):(0:0)):(((0:000)(00:0):(0:00)))	(((000:0):(00:0)))	(((0:0))	(())
6	((((00:0):((0:00)((0:00)):(((0:00):(0000000000)))	(((0:00):((0:0)))	(((0:0))	(())
7	((((000000:0))(0:00):(00:0)):(((0:00):((0:0000)))	(((00:0):((0:0)))	(((0:0))	(())
9	(((((0:0))((0:0):((0:0)))(((0:0)):((0:0))):(((0:0000)):((0:000)((0:0))((00:000000)))	(((00:0))((0:0):((0:000)))	(((00):0)))	(())
10	(((((0:0))(0000:0)):((0:00)((0:0)):(((00:0):(00:00)(00:0)))	(((00:00):((0:00)))	(((0:0))	(())
	→ →			
	Simpler Time-Series Representation			

The table is organized by number of iterations: 3 for 9 participants and 4 for 5 participants. As iteration level increases, the representation appears simpler.
doi:10.1371/journal.pone.0104087.t002

Table 3. Accumulation and dissipation rates computed from the Level 1 iteration for all participants.

	Accumulation Rates			Dissipation Rates		
Participant	**Mean (95% Confidence)**		**R^2**	**Mean (95% Confidence)**		**R^2**
ID	**μg/(dL • min)**			**μg/(dL • min)**		
1	0.18	(0.11, 0.26)	0.41	0.06	(0.03, 0.09)	0.28
2	0.15	(0.11, 0.20)	0.58	0.05	(0.03, 0.07)	0.49
3	0.14	(0.06, 0.22)	0.27	0.08	(0.04, 0.12)	0.35
4	0.25	(0.18, 0.32)	0.59	0.10	(0.08, 0.12)	0.64
5	0.24	(0.18, 0.29)	0.74	0.10	(0.06, 0.13)	0.53
6	0.32	(0.25, 0.39)	0.78	0.10	(0.07, 0.12)	0.71
7	0.21	(0.15, 0.27)	0.67	0.08	(0.05, 0.11)	0.56
8	0.13	(0.07, 0.19)	0.43	0.10	(0.08, 0.13)	0.73
9	0.17	(0.15, 0.39)	0.73	0.08	(0.05, 0.11)	0.52
10	0.17	(0.09, 0.26)	0.45	0.07	(0.03, 0.11)	0.39
11	0.27	(0.18, 0.36)	0.54	0.06	(0.03, 0.09)	0.42
12	0.33	(0.27, 0.39)	0.84	0.09	(0.05, 0.19)	0.54
13	0.07	(−0.12, 0.15)	0.14	0.05	(0.02, 0.09)	0.34
14	0.11	(0.09, 0.13)	0.86	0.06	(0.03, 0.09)	0.43
mean: 0.20 μg/(dL • min)			*mean:* 0.08 μg/(dL • min)			
minimum: 0.07 μg/(dL • min)			*minimum:* 0.05 μg/(dL • min)			
maximum: 0.33 μg/(dL • min)			*maximum:* 0.10 μg/(dL • min)			

doi:10.1371/journal.pone.0104087.t003

($R^2 = 0.78$). Fitting a regression line to the falling points results in a 0.098 μg/(dL min) dissipation rate ($R^2 = 0.71$). . Fitting regression lines allows computing average accumulation and dissipation rates for the time-series. The slope of the regression is analogous to secretion and clearance rates.

Multi-Scale Analysis. Using equations 17–19 the accumulation rate, dissipation rate, and inter-nadir interval at each iteration for participant ID6 are shown in **Figure 5 A–C**. At each iteration, the accumulation and dissipation rates and the point spread decrease. This decrease in point deviation is due to the change in time-scale at each iteration. The fastest accumulation rates occur at the end of scheduled sleep time and many of the fastest dissipation rates occur during sleep. The dissipation-accumulation plot (**Figure 5D**) shows an approximately log-log relationship between accumulation and dissipation across alliterations. The multi-scale analysis provides a way to assess the effect of a sequence of pulses on concentration profiles.

Accumulation Rates and Dissipation Rates. The average accumulation rate for the 14 participants was 0.20 μg/(dL min) and ranged between 0.07 and 0.33 μg/(dL min) when all the points from the first iteration are included (**Table 3**). The R^2 value for each participant, as computed during the linear regression of rises and falls in the data, ranged between 0.14 and 0.86. The average dissipation rate was 0.08 μg/(dL min) and ranged between 0.05 and 0.10 μg/(dL min); its R^2 value ranged between 0.28 and 0.73. There was a positive correlation between the accumulation and dissipation rate (Pearson = 0.44).

Multi-Scale Analysis. HAP analyses produces results on at least two time scales: the pulse scale in the order of minutes and the ultradian scale (less than 24 hours). For the 9 participants that converged in 3 iterations, the pulse scale and ultradian scale components are, 0.67 and 2.67 hours, respectively. In the 5 participants that converged in 4 iterations, HAP computed a pulse scale and two ultradian scale components corresponding to 0.67, 2.33, and 8.17 hours. We postulate that the participants whose hormone time series converge in 3 vs. 4 iterations may have differences in cortisol control. These inter-participant differences appear at the second iteration index in the pulsicons (**Table 2**); pulse clustering (i.e., start of a new pulse before the effects of previous pulses are completely dissipated) occurs in those who converge in 4 iterations. Further details of how pulsicons represent these differences are included in **Text S1**.

Discussion

HAP analysis demonstrates the applicability of CFLs to biological time-series data. This CFL for representing pulsatile time-series is novel and complementary to CFL methods developed for other fields [45–49]. HAP analysis is an objectively reproducible time domain method that efficiently characterizes the visually salient features in the data. Additional advantages are that HAP requires a minimum set of assumptions and does not require the user to set analysis parameters such as range or threshold. HAP analysis partitions the time-series temporally into hierarchically organized segments that may facilitate insight into the individual and group nature of pulsatility. The CFL provides a seamless interface for subsequent multi-scale analysis. HAP is efficient, providing individualized estimates of parameters in seconds of computing time. When applied to pulsatile time-series data, HAP provides quantitative estimates analogous to pharmacokinetic parameters and qualitative information through the use of production graphs and the novel text representation called pulsicons.

Novel concepts introduced

Language-Based Pulse Analysis Technique and Analysis Pipeline. The ability to automatically describe pulsatility formally is a novel HAP analysis feature. This feature facilitates

Table 4. Hierarchical Features Summary by Iteration Level for a single participant (ID6).

	Iteration									
	Level 1			Level 2			Level 3			Level 4*
	Min.	Median	Max.	Min.	Median	Max.	Min.	Median	Max.	Median
Rise duration (minutes)	10	10	40	30	60	140	90	160	210	620
Fall duration (minutes)	10	30	110	40	130	400	410	420	430	520
Inter-nadir interval (minutes)	20	40	150	90	185	430	520	570	620	1140
Rise amplitude (μg/dl)	0.04	0.93	10.98	0.92	4.66	8.57	2.57	5.97	8.04	3.09
Fall amplitude fall (μg/dl)	0.01	1.22	11.86	0.14	2.5	8.57	4.96	5.38	5.81	3.24

*Level 4 is the terminal iteration for participant ID6. Since there are only two points at the terminal iteration, only the median (defined for two points as the average) is shown.

doi:10.1371/journal.pone.0104087.t004

the development of an analysis pipeline, and it is intrinsically tied to visualizations. This theoretical framework for HAP is implemented as an extendable analysis pipeline whose design is motivated by automated methods derived from CFLs and current hormone analysis programs [15,16] and with the aim of translating HAP methods into use. Our pipeline provides support for analyzing data collected in a database, which differs from many current interfaces that analyze data from one participant at a time [15,16].

Hierarchically Organized Results. The hierarchically organized results from HAP analysis motivated a system of visualizations for summarizing pulsatility present in a time-series. For example, plots that illustrate the multi-scale nature of pulsatility (e.g., **Figure 5**) both provide a way to evaluate how features and parameters of the system are changing at different time-scales and provide insight into the number and organization of pulses. Visualizations for identifying the embedded nature of cortisol pulsatility (e.g., **Figure 3**) are another novel application of CFL production graphs. Pulsicons provide a new tool for understanding the pulsatility embedded within a hormone time-series through the ability to view the time-series as a text string. The pulsicons provide a framework to characterize and to potentially search time-series; this is a topic for future work. Linking HAP analysis output to visualizations is designed to facilitate verifying, validating, and analyzing results and to empower the user to explore features present in their data. The interface between generating HAP results and generating graphs is designed to employ grammar generated visualizations [50], which will allow for new visualizations to be generated interactively by the user.

Implications of cortisol pulsatility results

The HAP analysis presented in this paper demonstrates that the approach can identify known quantitative and qualitative cortisol pulsatility features [13,51,52]. (i) Mortlola et. al presented pulse secretion durations and amplitudes from women that are consistent with analogous values presented in this paper [51]. (ii) HAP analysis identified the same range of ultradian components of a 24 hour cortisol time-series as reported in Veldhuis et. al [13]. (ii) Our result of cortisol concentration levels in healthy participants rising during sleep, spiking to highest levels upon wake time, and falling to lowest levels near bedtime or early in the sleep episode is also consistent with findings from other methods. This quiescent period is seen in healthy individuals, and its absence is associated with depression in women [51].These classic features of cortisol pulsatility can be identified in both the production graph and the pulsicon representation of the time-series. The quiescent period, where there are few or no pulses, in the later day is clearly identified in the production graph representation and the associated pulsicon (section VI of **Figure 3**).

Novel HAP Generated Cortisol Findings

The qualitative and quantitative HAP analysis presented in this paper supports a complex view of cortisol pulsatility. The production graphs and pulsicons for each individual suggest a range of patterns in cortisol pulsatility over a 24-hour period that may reflect individual differences, different physiologic states (e.g., sleep or stress), pharmacology and/or response to environmental exposures. This suggests that methods for analyzing cortisol may need to adapt to individual differences present in the time-series. Specific HAP-identified findings are described below.

Expression of an Eight Hour Ultradian Rhythm in a Subset of Participants. HAP identified an 8-hour periodicity in the subset of participants that converged in 4 iterations and that

was not present in the group that converged in 3 iterations. The approximate 0.7 and 2.5 hour signals were identified in all participants by HAP; Veldhuis identified similar periodicities in 24-hour cortisol concentration profiles [13]. However, Veldhuis did not report an 8 hour periodicity (See **Multi-scale Analysis** section in **Group Results**). We conjecture that the inter-individual difference in ultradian signal could be due to individual differences in how time-of-day affects the Hypothalamic-Pituitary-Adrenal (HPA) axis or in the effect of the 8-hour sleep episode on cortisol pulsatility. Our conjecture that sleep is implicated is supported by a recent article that reports ultradian components in rodent sleep [53].Future work including more than 24 hours of data and more than one sleep episode per person will be needed to define the characteristics of this finding.

Qualitative Structure of Cortisol. In this paper, we report an objectively derived qualitative structure of cortisol pulsatility that is applicable to a range of cortisol profiles and is represented visually as a production graph and text string. These qualitative structures are representation patterns of pulsatility. Since patterns of hormone pulsatility affect responsiveness to the hormone [54–56], it is important to understand and define the underlying patterns. We hypothesized that the qualitative structure of cortisol pulsatility reflects the effect of sleep on cortisol production and is associated with ultradian signals.

Method Limitations

The formal pulsatility definition in this method requires an extension to the traditional definitions of pulsatility to include multiple time-scales and the pharmacokinetic measures of accumulation, dissipation, and inter-nadir interval. Because the method does not distinguish between 'pulse counting' and noise within the time-series, the results from the output of the first iteration may include both pulses and noise. Thus, the results from the first iteration are individualized lower and upper bound of values that include noise. Although noise, such as assay noise, may not be physiologically meaningful, the inclusion of low amplitude events within the results does have advantages. For example, the presence of a significant number of low amplitude events may suggest increased negative feedback, suppression of the firing signal, or sensitivity changes in the system's response to input, either hypothalamic or from the pituitary gland. The large quantity of low amplitude events within the cortisol data suggests that noise is an important feature of the system; a challenge is that noise characteristics may change with time of day or disease state. Traditional 'smoothing' of the data may eliminate or minimize noise may destroy information required to understand HPA axis control. So as to better differentiate between assay and system noise, future versions of HAP could include an assay error term or other error term suitable for the signal under study.

Future work

Further research is required to link hierarchically organized segments with mathematical statements of relevant physiology. Linking HAP to physiological concepts through the use of observational dynamical models [12,13,24,33,57] and cellular models could provide an approach to testing relationships between HAP partitions and mechanisms of HPA control. For example, observational dynamical models could link ACTH, which stimulates cortisol release, with cortisol time-series. These models could be used to estimate higher order pharmacokinetic terms as well as to test specific hypotheses about different elements within a hierarchical control system that includes feedback as well as other elements, such as circadian rhythms or sleep-wake state, that affect their production. Application of HAP to different physiologic and pathophysiologic cortisol time-series data, including segments with and without sleep, will reveal how HAP can be used to quantitatively and qualitatively explore changes associated with these conditions [13,58–68].

Supporting Information

Figure S1 Specification of a CFL for recognizing parenthesis.
(TIF)

Figure S2 Processing of a string of parenthesis with a CFL.
(TIF)

Figure S3 HAP_Analysis output.
(TIF)

Figure S4 Definition and interpretation of the comparator operator.
(TIF)

Figure S5 Linking time-series (data), the production graph and pulsicons.
(TIF)

Text S1 Supportive text and examples.
(DOCX)

Text S2 Includes background and additional examples.
Table S1. Fields description for HAP_Analysis input structure.
Table S2. Fields description for HAP_Analysis Program Data Structure.
(DOCX)

Acknowledgments

We thank Drs. Kun Hu, Janis Anderson, and Kimberly Rapp for discussing the relevance of the presented research to general hormone analysis methods, clinical assessment, and animal studies, respectively. Dr. Kimberly Rapp provided detailed comments on an early draft of this manuscript that guided subsequent revisions. Dr. Joseph Hull and Jason Sullivan made many ongoing suggestions for improving the presentation of these methods. We would also like to thank Journal editors and blind reviewers who provided specific recommendations which guided refinement of the manuscript.

Author Contributions

Conceived and designed the experiments: DAD. Performed the experiments: DAD. Analyzed the data: DAD. Contributed reagents/materials/analysis tools: DAD EBK GKA DPN. Wrote the paper: DAD EBK GKA DPN.

References

1. Haefner JW (2008) Modeling Biological Systems Principles and Applications. New York: Springer. 1–475 p.
2. Clifton DK (1987) Objective validation of pulse detection methods by computer simulation. In: Crowley WF Jr., Hofler JG, editors. The episodic secretion of hormones. New York: John Wiley & Sons. pp. 35–46.
3. Lambalk CB, de Koning J, van Kessel H, van Rees PG, Schoemaker J (1987) Calculation of the intra-assay variation per assay and its relevance to leuteinizing hormone pulse detection. In: Crowley WF Jr., Hofler JG, editors. The episodic secretion of hormones. New York: John Wiley & Sons. pp. 67–117.
4. Merriam GR, Wachter KW (1982) Algorithmns for the study of episodic hormone secretion. American Journal of Physiology: Endocrinology and metabolism 243: E310–E318.

5. Keenan DM, Veldhuis JD (1998) A biomathematical model of time-delayed feedback in the human male hypothalamic-pituitary-Leydig cell axis. Am J Physiol 275: E157–E176.
6. Van Cauter E (1974) Methods for the analysis of multifrequential biological time series. JInterdisciplCycle Res 5: 131–148.
7. O'Sullivan F, O'Sullivan J (2007) Deconvolution of Episodic Hormone Data: An analysis of the Role of Season on the Onset of Puberty in Cows. Biometrics 44: 339–353.
8. Urban RJ, Kaiser DL, Van Cauter E, Johnson MJ, Veldhuis JD (1988) Comparative assessments of objective peak-detection algorithms. II. Studies in men. American Journal of Physiology - Endocrinology and Metabolism 254: E113–E119.
9. Veldhuis JD, Johnson ML (1986) Cluster analysis: A simple, versatile, and robust algorithm for endocrine pulse detection. AmJPhysiol 250: E486–493.
10. Naska A, Oikonomou E, Trichopoulou A, Psaltopoulou T, Trichopoulos D (2007) Siesta in healthy adults and coronary mortality in the general population. Arch Intern Med 167: 296–301.
11. Veldhuis JD, Johnson ML (1994) Analytical methods for evaluating episodic secretory activity within neuroendocrine axes. NeurosciBiobehavRev 18: 605–612.
12. Jusko WJ, Slaunwhite WR Jr., Aceto T Jr. (1975) Partial pharmacodymanic model for the circadian-episodic secretion of cortisol in man. JClinEndocrinolMetab 40: 278–289.
13. Veldhuis JD, Johnson ML, Iranmanesh A, Lizarrallde G (1990) Temporal Structure of In Vivo Adrenal Secretory Activity Estimated by Deconvolution Analysis. Journal of Biological Rhythms 5: 247–255.
14. Santen RJ, Bardin CW (1973) Episodic luteinizing hormone secretion in man. JClinInvest 52: 2617–2628.
15. Van Cauter E, Hasak M, Leproult R, Silbersweig D (2004) Chronobiological Series Analyzer. 1.0 ed.
16. Johnson ML, Pipes L, Veldhuis PP, Farhy LS, Boyd DG, et al. (2008) AutoDecon, a deconvolution algorithmn for identification and characterization of lutenizing hormone secretory bursts: Description and validation using synthetic data. Analytical Biochemstry 381: 8–17.
17. Brown EM, Enyedi P, Hawkins D, Williams G (1988) Binding sites for inositol trisphosphate (IP3) and tetrakisphosphate (IP4) in parathyroid microsomes. < None Specified>: S75–S75.
18. Brown EN, Choe Y, Luithardt H, Czeisler CA (2000) A statistical model of the human core-temperature circadian rhythm. Am J Physiol Endocrinol Metab 279: E669–E683.
19. Klerman EB, Adler GK, Jin M, Maliszewski AM, Brown EN (2003) A statistical model of diurnal variation in human growth hormone. Am J Physiol 285: E1118–E1126.
20. Brown EN, Meehan PM, Dempster AP (2001) A stochastic differential equation model of diurnal cortisol patterns. AmJPhysiolEndocrinolMetab 280: E450–E461.
21. Keenan DM, Alexander S, Irvine C, Veldhuis JD (2009) Quantifying Nonlinear Interactions within the Hypothalmo-Pituatary-Adrenal Axis in the Conscious Horse. Endocrinology 150: 1941–1951.
22. Keenan DM, Alexander S, Irvine C, Veldhuis JD (2009) Quantifying nonlinear interactions within the hypothalamo-pituitary-adrenal axis in the conscious horse. Endocrinology 150: 1941–1951.
23. Gupta S, Aslakson R, Gurbaxani BM, Vernon SD (2007) Inclusion of the glucocorticoid receptor in a hypothalamic pituitary adrenal axis model reveals bistability. Theoretical Biology and Medical Modelling 4: 1–12.
24. Walker JJ, Terry JR, Lightman SL (2010) Origin of ultradian pulsatility in the hypothalamic-pituitary-adrenal axis. Proceedings of The Royal Society B 277: 1627–1633.
25. Veldhuis JD, Keenan DM, Pincus SM (2008) Motivations and Methods for Analyzing Pulsatile Hormone Secretion. Endocrine Reviews 29: 823–864.
26. Du D-Z, Ko K-I (2001) Problem Solving in Automata, Languages, and Complexity: John Wiley & Sons, Inc. 1–396 p.
27. Sipser M (1997) Introduction to the Theory of Computation: Course Technology. 1–416 p.
28. Lewis RD, Warman GR, Saunders DS (1997) Simulations of free-running rhythms, light entrainment and the light-pulse phase response curves for the locomotor activity rhythm in *period* mutant of *Drosphila melanogaster*. JTheorBiol 185: 503–510.
29. Pullum G, Gazdar G (1982) Natural languages and context-free languages. Linguistics and Philosophy 4: 471–504.
30. Chomsky N (1956) Three models for the description of language. IRE Transaction on Information Theory 2: 113–124.
31. Knuth DE (1968) Semantics of Context-Free Languages. Mathematical Systems Theory 2: 127–145.
32. United States Food and Drug Administration (FDA) (2007) Guidance for Industry: Computerized Systems Used in Clinical Investigation. 1–13 p.
33. Brown EN, Meehan PM, Demster AP (2001) A stochastic differential equation model of diurnal cortisol patterns. American Journal of Physiology - Endocrinology and Metabolism 280: E450–E461.
34. Klerman EB, Goldenberg DL, Brown EN, Maliszewski AM, Adler GK (2001) Circadian rhythms of women with fibromyalgia. J Clin Endocrinol Metab 86: 1034–1039.
35. Duffy JF, Wright Jr KP (2005) Entrainment of the human circadian system by light. J Biol Rhythms 20: 326–338.
36. Minors DS, Waterhouse JM (1984) The use of constant routines in unmasking the endogenous component of human circadian rhythms. ChronobiolInt 1: 205–216.
37. Mills JN, Minors DS, Waterhouse JM (1974) The circadian rhythms of human subjects without timepieces or indication of the alternation of day and night. JPhysiol(Lond) 240: 567–594.
38. Lewis HR, Papadimitriou C (1998) Context Free Languages. Elements of the Theory of Computation: Prentice Hall. pp. 113–178.
39. Proschak E, Wegner JK, Shüller A, Schneider G, Fechner U (2010) Molecular Query Language (MGL) - A Context-Free Grammar for Substructure Matching. Journal of Chemical Information and Modeling 47: 295–301.
40. Phaninadra G, Shankar KR, Sreenivas PD (2007) A Fast Multiple Matching Algorithmn using Context Free Gammar and Tree Model. International Journal of Computer Science and Network Security 7: 231–234.
41. Zarchan P, Musoff H (2011) Fundamentals of Kalman Filtering: A Practical Approach; Frank KL, editor. Reston, Virginia: American Institute of Aeronautics and Astronautics. 1–852 p.
42. Barrett CL, Reidys CM (1999) Elements of a theory of computer simulation I: Sequential CA over random graphs. Applied Mathematics and Computation 98: 241–259.
43. Barrett CL, Hunt Iii HB, Marathe MV, Ravi SS, Rosenkrantz DJ, et al. (2003) Predecessor and Permutation Existence Problems for Sequential Dynamical Systems. Discrete Mathematics and Theoretical Computer Science AB: 69–80.
44. Mortveit HS, Reidys CM (2008) An Introduction to Sequential Dynamical Systems: Springer Science + Buisness Media. 1–248 p.
45. Wehr TA (2001) Photoperiodism in humans and other primates: Evidence and implications. Journal of Biological Rhythms 16: 348–364.
46. Wang D, Teichtahl H, Goodman C, Drummer O, Grunstein RR, et al. (2008) Subjective daytime sleepiness and daytime function in patients on stable methadone maintenance treatment: possible mechanisms. J Clin Sleep Med 4: 557–562.
47. Magalooikonomou V, Wang W, Li G, Faloutsos C (2005) A Multiresolution Symbolic Representation of Time Series. pp. 668–679.
48. Wang W, Portnoy V, Pollak I (2007) A Stochastic Context-Free Grammar Model for Time Series Analsysis. Available: http://dx.doi.org/10.1109/ICASSP.2007.367069
49. Almajai I, Milner B, Darch J, Vaseghi S (2007) Visually-Derived Wiener Filters for Speech Enhancment. IEEE International Conference on Acoustics, Speech and Signal Processing Proceedings: IV-585–IV-588.
50. Wilkinson L (2005) The Grammar of Graphics: Springer. 1–694 p.
51. Mortola JF, Liu JH, Gillin C, Rasmussen DD, Yen SSC (1987) Pulsatile Rhythms of Adrenocorticotropin (ACTH) and Cortisol in Women with Endogenous Depression: Evidence for Increased ACTH Pulse Frequency. Journal of Clinical Endocrinology and Metabolism 65: 962–968.
52. Stewart PM (2008) The Adrenal Cortex. In: Kronnenbert HM, Melmed S, Polonsky KS, Larsen PR, editors. Williams Textbook of Endocrinology. Philadelphia: W.B. Sunders Company. pp. 445–537.
53. Stephenson R, Lim J, Famina S, Caron AM, Dowse HB (2012) Sleep-Wake Behavior in the Rat: Ultradian Rythms in a Light-Dark Cycle and Continuous Bright Light. Journal of Biological Rhythms 27: 490–501.
54. Keller-Wood M (1998) ACTH responses to CRF and AVP in pregnant and nonpregnant ewes. AJP - Regulatory, Integrative and Comparative Physiology 274: R1762–R1768.
55. Matthews SG, Allis JRJ (1995) Regulation of CRH and AVP mRNA in the developing ovine hypothalamus: effects of stress and gluccocorticoids. American Journal of Physiology - Endocrinology and Metabolism 268: E1096–E1107.
56. Mattews SG, Challis JRG (2009) CRH and AVP-Induced Changes in Synthesis and Release of ACTH from the Ovine Fetal Pituitary In Vitro: Negative Influences of Cortisol. Endocrine 6: 293–300.
57. Acherman P (2004) The Two-Process Model of Sleep Regulation Revisited. Aviation, Space, and Environmental Medicine 75: A37–A43.
58. Beck-Friis J, Ljunggren JG, Thoren M, von Rosen D, Kjellman BF, et al. (1985) Melatonin, Cortisol and ACTH in Patients with Major Depressive Disorder and Healthy Humans with Special Reference to the Outcome of the Dexamethasone Supression Test. Psychoneuroendocrinology 10: 173–186.
59. Rubin RT, Phillips JJ, McCracken JT, Sadow TF (1996) Adrenal Gland Volume in Major Depression: Relationship to Basal and Stimulated Pituitary-Adrenal Cortical Axis Function. Biological Psychiatry 40: 89–97.
60. Koenigsberg HW, Teicher MH, Mitropoulou V, Navalta C, New AN, et al. (2009) 24-h Monitoring of plasma norepinephrine, MHPG, cortisol, growth hormone and prolactin in depression. Journal of Psychiatric Research 38: 503–511.
61. Pollmacher T, Mullington J, Korth C, Schreiber W, Hermann D, et al. (1996) Diurnal Variations in the Human Host Response to Endotoxin. The Journal of Infectious Diseases 175: 1040–1045.
62. Mullington JM, Haack M, Toth M, Serrador JM, Meirer-Ewert HK (2009) Cardiovascular, Inflamatory, and Metabolic Consequences of Sleep Deprivation. Progress in Cardiovascular Diseases 51: 294–302.
63. Buckly TM, Schatzberg AF (2005) On the Interactions of the Hypothalamic-Pituitary-Adrenal (HPA) Axis and Sleep: Normal HPA Axis Activity and Circadian Rhythm, Exemplary Sleep Disorders. The Journal of Clinical Endocrinology & Metabolism 90: 3106–3114.
64. De Kloet ER (2010) About Stress hormones and Resilience to Psychopathology. Journal of Neuroendocrinology 20: 885–892.

65. McEwen BS (2007) Physiology and Neurobiology of Stress and Adaptation Central Role of the Brain. Physiology Review 87: 873–904.
66. Sharpe M, Chalder T, Palmer I, Wessely S (2010) Chronic Fatigue Syndrome: A Practical Guide to Assesment and Management. General hospital Psychiatry 19: 185–199.
67. Lopowski ZJ (2010) Chronic Idopathic Pain Syndrome. Annals of Medicine 22: 213–217.
68. Schweinhardt P, Sauro KM, Bushnell MC (2008) Fibromyalgia: A Disorder of the Brain. Neuroscientist 16: 415–421.