Biophysical Chemistry 152 (2010) 184–190

Contents lists available at ScienceDirect

Biophysical Chemistry

journal homepage: http://www.elsevier.com/locate/biophyschem

Statistical analysis of plasma thermograms measured by differential scanning calorimetry

Daniel J. Fish [a,e,*], Greg P. Brewood [a,e], Jong Sung Kim [b], Nichola C. Garbett [c], Jonathan B. Chaires [c], Albert S. Benight [a,d,e]

[a] Portland Bioscience, Inc., Portland, OR, United States
[b] Fariborz Maseeh Department of Mathematics and Statistics, Portland State University, Portland, OR, United States
[c] James Graham Brown Cancer Center, University of Louisville, Louisville, KY, United States
[d] Departments of Chemistry and Physics, Portland State University, Portland, OR, United States
[e] Louisville Bioscience, Inc., Louisville, KY, United States

ARTICLE INFO

Article history:
Received 30 July 2010
Received in revised form 22 September 2010
Accepted 22 September 2010
Available online 29 September 2010

Keywords:
Plasma thermograms
Diagnostics
Calorimetry
Biostatistics
Chemometrics

ABSTRACT

Melting curves of human plasma measured by differential scanning calorimetry (DSC), known as thermograms, have the potential to markedly impact diagnosis of human diseases. A general statistical methodology is developed to analyze and classify DSC thermograms to analyze and classify thermograms. Analysis of an acquired thermogram involves comparison with a database of empirical reference thermograms from clinically characterized diseases. Two parameters, a distance metric, P, and correlation coefficient, r, are combined to produce a 'similarity metric,' ρ, which can be used to classify unknown thermograms into pre-characterized categories. Simulated thermograms known to lie within or fall outside of the 90% quantile range around a median reference are also analyzed. Results verify the utility of the methods and establish the apparent dynamic range of the metric ρ. Methods are then applied to data obtained from a collection of plasma samples from patients clinically diagnosed with SLE (lupus). High correspondence is found between curve shapes and values of the metric ρ. In a final application, an elementary classification rule is implemented to successfully analyze and classify unlabeled thermograms. These methods constitute a set of powerful yet easy to implement tools for quantitative classification, analysis and interpretation of DSC plasma melting curves.

1. Introduction

Analytical methods for proteomics are used to survey the protein content of biological fluids and cells by monitoring relative amounts of various constituent proteins in the sample mixture. Advances in mass spectrometry, computational methods, automated software routines, and data storage have paved the way for development of increasingly sophisticated methods for proteomic analysis. However, applying such methods for diagnostic detection requires accurate identification of disease states based on deviations from cellular homeostasis. For instance, disease states can manifest as slight disturbances of protein concentration levels and/or intermolecular interactions among different proteins and other components. Even without a clear understanding of underlying cellular processes that give rise to these perturbations and how they are correlated with disease, complex patterns derived from proteomic analysis of biological fluids can provide accurate and reliable identification of different disease states [1–9]. Such diagnostic capability has the potential to empower more effective treatment programs during the early stages of disease development, when aggressive actions may be most critical and provide the greatest health benefit.

Recently, Chaires and co-workers have reported on the remarkable utility of differential scanning calorimetry (DSC) as a diagnostic tool for analysis of complex biological mixtures such as plasma from human blood [1–4]. DSC measures the heat change of a sample as it is heated over a temperature range from ~293 K to 400 K. Increasing the temperature of the biological fluid induces structural transitions in certain molecular constituents present in the fluid (e.g. proteins). Generally, a release (exothermic reaction) or absorption (endothermic reaction) of small amounts of heat accompanies these transitions. A DSC melting profile, or thermogram, is a plot of the excess specific heat capacity as a function of temperature (T).

Published results have demonstrated the extraordinary potential of DSC analysis to differentially distinguish human plasma samples from healthy (i.e. clinically non-symptomatic) individuals and patients suffering from a variety of clinically diagnosed diseases, specifically, various autoimmune diseases and different types of cancer [1–4]. It has been shown that plasma thermograms of samples from clinically diagnosed patients with rheumatoid arthritis (RA), systemic lupus erythematosus (lupus) and Lyme disease are dramatically different

* Corresponding author. Tel.: +1 503 445 8121; fax: +1 503 473 8379.
E-mail address: djf@pdxbio.com (D.J. Fish).

0301-4622/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2010.09.007

from one another and from those derived from samples of clinically diagnosed 'healthy' patients [1–4].

For this study, thermograms of over 2000 plasma samples, representing at least 25 different disease states, have been recorded for analysis. Overall, these patterns have proven to be remarkably consistent and characteristic. In many cases, differences in curve shape are easily distinguished but, for other cases, subtle differences must be extracted using quantitative statistical methods. To perform this analysis a unique, nonparametric, quantitative and reliable statistical scheme was developed to compare and classify thermograms according to their shapes and intrinsic patterns.

The statistical methods developed and tested here perform several key functions: (1) statistical characterization of large numbers of similar DSC thermograms for use as comparative reference sets; (2) comparison of acquired thermograms (herein called test curves) to sets of reference curves and quantitative determination of degree of similarity or difference between test and reference curves, and (3) classification of a test thermogram according to disease state.

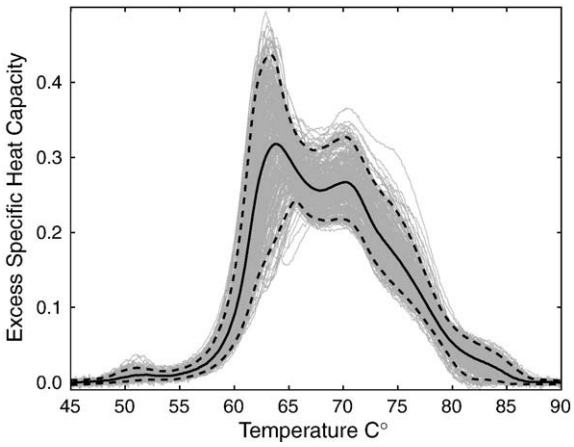

Fig. 1. Plot of all plasma thermograms collected from over 500 healthy subjects shown with median thermogram (heavy solid curve) and 90% quantiles (dashed curves). Plasma samples were in a buffer containing 10 mM potassium phosphate, 150 mM NaCl, 0.38% (w/v) sodium citrate, pH 7.5, and contained approximately 2 mg/mL total protein. A scan rate of 60 deg/hr was used, and data were collected over a range of 20–110°.

2. Materials and methods

2.1. Overview

Analytical methods were developed to enable identification of unclassified thermograms through quantitative comparisons with well-categorized reference sets of known healthy and disease thermograms. First, statistical characterization of reference thermograms defines basic parameters required to establish a template for a particular disease category. Next, a similarity metric is employed to determine the extent to which unclassified thermograms are similar to a given reference set.

For the purposes of this development, the extent to which a particular set of reference thermograms completely or adequately represents the entire population afflicted with a particular disease is not evaluated. To avoid statistical issues with estimation of population parameters and questions of sample size, we assume throughout that all reference sets are sufficiently representative of the population in question. Further analysis to estimate the population distribution of DSC thermograms for various disease states is currently in progress through clinical studies of thermograms from large numbers of plasma samples.

2.2. Thermograms of plasma samples

All thermograms used in this study were obtained at the Biophysics Core Facility at the University of Louisville. A total of 600 thermograms from blood plasma samples received from the Lupus Family Registration and Repository (Oklahoma City, OK) were collected from 300 patients clinically diagnosed with systemic lupus erythematosus (lupus). Additionally, over 500 blood plasma samples were collected from patients clinically determined to be healthy individuals (displayed in Fig. 1). Thermograms from this set were used as 'Healthy' reference data for analysis and comparison.

The effect of scan rate on plasma thermograms was explored by DSC experiments using scan rates of 60°, 120° and 200°/h. Faster scan rates resulted in a slight elevation of melting temperatures of at least some plasma components, but the overall shape of the thermogram was retained, and dramatic differences between normal controls and diseased plasma were still observed. We chose a standard operating protocol that used a scan rate of 60°/h, and all thermograms were collected using the identical protocol.

2.3. Characterization of reference sets

A DSC thermogram can be expressed as a n-dimensional vector $X(T) = (x_1, x_2, \ldots, x_n)$ with each entry x_i corresponding to the measured heat capacity at each temperature T_i. Data are typically collected over temperatures ranging from 25 to 110 °C, with 10 measurements for each degree (°C), resulting in $n = 850$. In practice, thermograms are often truncated to a range of 45 to 95°C ($n = 500$) since the most relevant thermodynamic information is usually contained in this region. A reference set of thermograms is then a collection of m vectors $\{X_j(T) | j = 1, \ldots, m\}$, where each $X_j(T)$ represents a single reference thermogram. The median thermogram, $X_{ref}(T)$, of a reference set is computed from all curves within the set and serves as a template thermogram representing that category. Variance within a reference set is quantified at each temperature by upper and lower quantile vectors, $\sigma_{upper}(T)$ and $\sigma_{lower}(T)$, respectively. These vectors establish the 0.05 to 0.95 quantile boundaries, wherein 90% of the measured data lies. In the case that a more detailed knowledge of the distributions underlying the data is obtained, these parameters can be refined using distribution-dependent measures of central tendency and variance.

The quality of a potential reference set of thermograms for a particular disease state is assessed using descriptive statistical measures. To determine the extent that a given curve in the reference set aligns with the median thermogram, each reference curve is compared to the median thermogram and the linear correlation coefficient (r value) is determined, resulting in a distribution of r values. High levels (>0.8) of correlation indicate that a given reference set is homogeneous in shape, and is taken as an indication that the median thermogram $X_{ref}(T)$ can be reliably used as a template curve. For simplicity, the Pearson's correlation coefficient is used here; however, more general and non-parametric correlations (such as Kendall's tau) could also be used when appropriate [10]. Quantile box-plots are also constructed to assess the variability or degree of homogeneity in a reference set.

2.4. Thermogram comparisons

Methods were aimed primarily at addressing the diagnostic need, i.e., determining to what degree an unclassified test curve, $X_{test}(T)$, aligns with a given reference template curve, $X_{ref}(T)$. The degree of similarity between a test curve and a reference thermogram can be characterized by two factors: (1) closeness in space (standardized Euclidean distance) at each temperature point, and (2) similarity in shape (correlation). In general, two thermograms can be highly correlated but, due to vertical scaling, may be separated by a nontrivial distance in space. Conversely, two thermograms can be

spatially close but poorly correlated due to fluctuations or noise in the data. For these reasons, the metric employed to quantify similarities between test and reference curves is a combination of both spatial distance and linear correlation.

Distance between two curves is determined using a similarity index, $P(X_{test}, X_{ref}, \sigma_{lower}, \sigma_{upper})$ defined as follows. At each temperature T, the standardized distance between X_{test} and X_{ref} is calculated as,

$$d(T) = abs\left[\left(X_{test}(T) - X_{ref}(T)\right) / \sigma_{ref}(T)\right],$$

where $\sigma_{ref}(T) = X_{ref}(T) - \sigma_{lower}(T)$ if $X_{test}(T) < X_{ref}(T)$ and $\sigma_{ref}(T) = X_{ref}(T) + \sigma_{upper}(T)$ if $X_{test}(T) > X_{ref}(T)$. The standardized distance $d(T)$ can be interpreted as the distance associated with a given reference data set that takes into account empirical distributions in the form of quantiles recorded for particular data sets (see Fig. 2). A value of $d(T) > 1$ indicates that, at the temperature T, the test curve is more distant from the median than 90% of the data in the reference set. At each temperature, define:

$$p(T) = \begin{cases} 0.9 & \text{if } d(T) \le 1 \\ 0.1 & \text{if } d(T) > 1 \end{cases}$$

The function $p(T)$ returns high values (0.9) at temperatures for which the test curve falls within the quantile boundary, and returns low values (0.1) at temperatures for which the test curve falls outside the quantile boundary. Thus, the function $p(T)$ represent a likelihood, based on quantile values, that the test curve is similar to the reference set at each temperature point. No assumptions are made about the distribution of the reference thermograms. As a result, this choice of function may not be optimal for discrimination of test and template thermograms. If more specific information about the distribution of reference curves is known, then σ_{ref} may be replaced with a distribution-based measure of variance, and a probability distribution function representing the reference set (e.g. Gaussian or logistic functions) can be used to compute critical values of $d(T)$ at each temperature T. In addition, an optimal temperature range may be empirically determined to enhance the discrimination between thermograms.

A scalar quantity representing similarity of the entire test curve to the reference set is then computed as the arithmetic mean of $p(T)$ over all temperatures $(T_i, i = 1, 2, \ldots, n)$:

$$P = \sum_{i=1}^{n} p(T_i) / n.$$

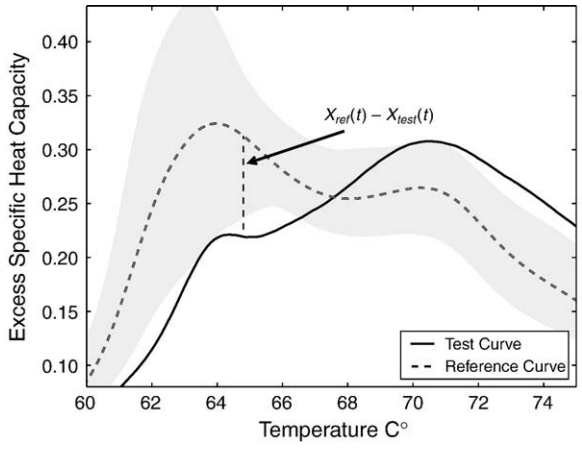

Fig. 2. Comparison of a test thermogram $X_{test}(T)$ (solid curve) to a reference median thermogram $X_{ref}(T)$ (dashed curve) with 5–95% quantile boundaries $\sigma_{ref}(T)$ (shading).

The metric P can be interpreted as a probability determined by the standardized multivariate distance between the test curve and the reference template. A value of P near unity indicates that the test curve is closer to the reference template than 90% of the reference data, while a value of P near zero indicates that the test curve is more distant than 90% of the data.

Similarity in shape between a test curve and the reference set is quantified using a linear correlation, r, such as Pearson's or Kendall's tau correlation [10]. Two thermograms that are linearly correlated necessarily possess similar shapes, so the linear correlation is an effective measure for discriminating between curves with different shapes (assuming similar scaling of the data). A linear correlation provides valuable information about the shape of test curves and helps to support and strengthen conclusions about degrees of similarity between test and reference curves. Due to similarities in the overall protein composition of human blood plasma, any two thermograms will be highly correlated in certain temperature regions. For instance, in very low (<50°C) and very high (>90°C) temperature regions major differences in thermogram shape are seldom found. As a result, the value of a linear correlation coefficient, r, between a test curve and a reference median curve will, in practice, never be negative, and will seldom even be close to zero. On an absolute scale, interpretation of r-values with regard to the strength of relationship must be done with some of care. In practice, initial characterizations of the data will help to determine significant levels of correlation (r) for interpretive use. However, for the purposes of comparing two thermograms, the relative scale of r is more valuable, and can be established from an empirical calibration using training data.

A composite parameter, ρ, is then defined as the weighted geometric average of the standardized distance metric, P, and the correlation coefficient, r. That is,

$$\rho = \left(P^w \cdot r^{2-w}\right)^{1/2},$$

where $0 \le w \le 2$ (in the unlikely case that $r \le 0$ we set $\rho = 0$). The weight w is typically selected through empirical optimization in order to maximize differences between distinct curve shapes. The range of ρ is [0,1], with values near zero indicating large differences in shape, and values near 1 indicating high similarity. The absolute scale for ρ depends on the particular reference set (or sets) employed, and at this stage has not been generally established. Instead, a relative scale is empirically determined based on training data, and the metric is calibrated before application to unknown test curves. The similarity metric, ρ, incorporates both distance and correlation into a single quantitative statistic that can then be used for discrimination between test curves and reference templates.

3. Results

To demonstrate application of our analytical methods, a set of 100 thermograms from samples clinically classified as 'Healthy' served as the Healthy reference set. Quantile boxes shown in Fig. 3 were constructed at a number of temperature points. Boxes represent 90% quantiles ranging from 5 to 95% while horizontal dashes represent median values. Small crosses indicate data points falling outside of the quantile range. The dashed curve is the median reference thermogram. For this set of thermograms, relatively little variance is observed in the lower (45–60 °C) and upper (80–90 °C) temperature ranges. In the central temperature region (60–80 °C), larger variability exists. Variations in the data can be visualized from a plot of the variance as a function of temperature. Fig. 4 shows that regions of low variance do indeed occur in the temperature ranges from 45–60 and 80–90 °C (average variance $s^2_{45-60} = 2.90 \times 10^{-5}$ and $s^2_{80-90} = 4.43 \times 10^{-5}$, respectively). Relatively higher variance is observed in the range from

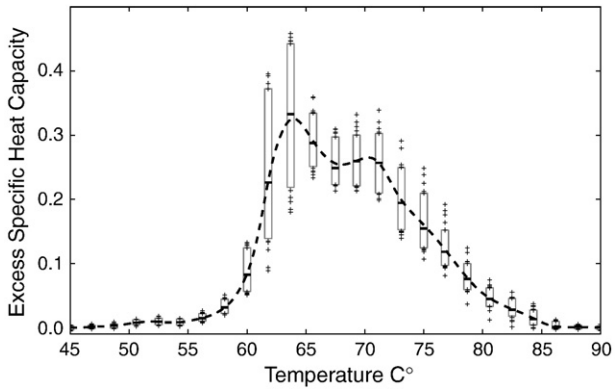

Fig. 3. Box plots of data from Healthy thermograms at various temperature points. Boxes represent the range between 5 and 95% quantiles. Horizontal lines represent median values at each temperature. Small crosses represent data points that fall outside of the quantile range. The mean thermogram is plotted as a dashed curve.

60 to 80 °C (average variance $s^2_{60-80} = 1.35 \times 10^{-3}$). The average variance over the entire temperature range was $s^2 = 6.0 \times 10^{-4}$.

To determine the degree to which reference thermograms align with the median Healthy thermogram, the linear correlation between each reference curve and the median thermogram was computed for each reference curve in the set (see Fig. 5). The average coefficient over all curves was $r_{avg} = 0.978 \pm 0.020$, which indicates very high levels of correlation between individual and median thermograms. This might reasonably be expected from a homogeneous population and lends support for using Healthy thermograms as a reference set when classifying unknown thermograms.

3.1. Test case — healthy versus lupus

As a test of the analytical scheme, one thermogram from a healthy patient withheld from the original reference set, along with one thermogram from a lupus patient was compared to the reference set of Healthy thermograms. As Fig. 6 shows, the test curve from the healthy patient falls entirely within the 90% quantile band surrounding the median reference Healthy thermogram. In fact, the largest distance between the sample curve and the reference thermogram was $d(T) = 1.10$, which occurred in the tail (below 50°C). The Lupus thermogram, on the other hand, fell significantly outside of the quantile band for a majority (>75%) of temperature points. Furthermore, the average distance from the Lupus test curve to the median Healthy curve ($d(T) = 2.76$) was larger than the maximum distance from the Healthy test curve to the median Healthy curve. The maximal

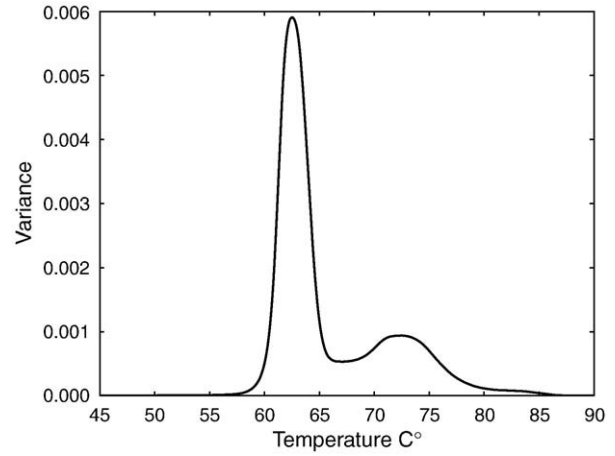

Fig. 4. Variances of Healthy thermograms plotted as a function of temperature.

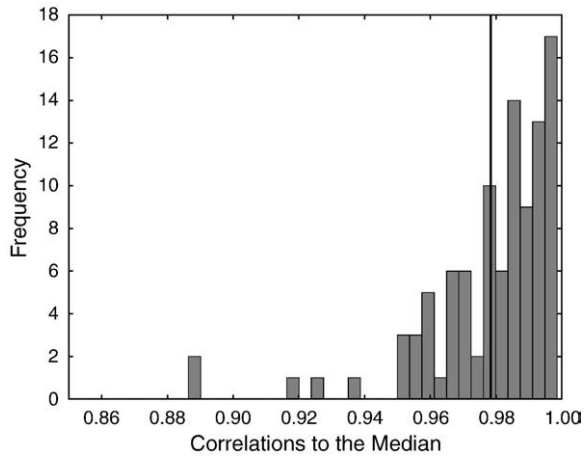

Fig. 5. Histogram of correlation coefficients between sample curves and the median Healthy thermogram. An average coefficient of $r_{avg} = 0.978$ (vertical line) demonstrates a high degree of correlation between sample curves and the median reference thermogram.

distance from the Lupus test curve to the median Healthy curve was $d(T) = 8.58$; significant differences are clearly evident.

Both graphical and numerical methods support the notion that the Healthy test curve can be classified as similar to the Healthy reference set while the Lupus test curve is classified as different. Quantitative evidence is provided by calculations of the similarity measures described above. With an optimal temperature range of 55–85 °C, and weight parameter $w = 1.0$, analysis yielded a value of $P = 0.868$ for the Healthy test curve (a high value) and $P = 0.289$ for the Lupus test curve (a low value). A correlation value of $r = 0.995$ was produced for the Healthy test curve and $r = 0.456$ for the disease test curve, and the composite ρ metric had values of $\rho = 0.929$ for the Healthy curve and $\rho = 0.363$ for the Lupus curve. Based on these methods, one can conclude that the two test curves represent distinct disease states.

3.2. Ideal case — thermogram simulations

To estimate the accuracy of these methods, simulations of plasma thermograms were performed. As a first approximation, thermograms were decomposed into weighted sums of Gaussian functions, which represent ideal melting profiles of generic proteins. The amplitude of each Gaussian peak is proportional to the protein's concentration in solution and the location of the peak along the abscissa indicates its melting temperature. To simulate subtle changes to thermograms that

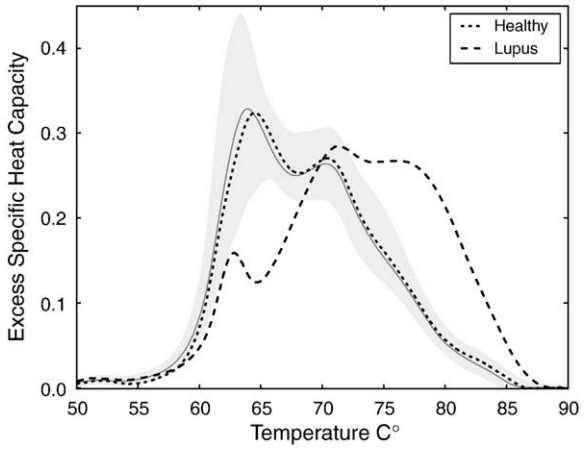

Fig. 6. Comparison of Healthy (dotted curve) and Lupus (dashed curve) test thermograms to the reference median thermogram (solid gray curve). Shaded region represents the 90% quantile band around the median.

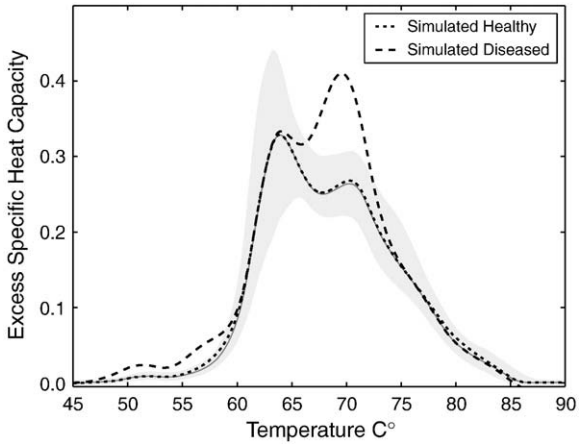

Fig. 7. Examples of simulated thermograms generated by the sum of the Healthy reference thermogram and randomly generated Gaussians. A 'simulated Healthy' thermogram with component protein concentration within the 90% quantile range is shown as a dotted curve. A 'simulated Diseased' thermogram with component protein concentration outside of the 90% quantile range is shown as a dashed curve.

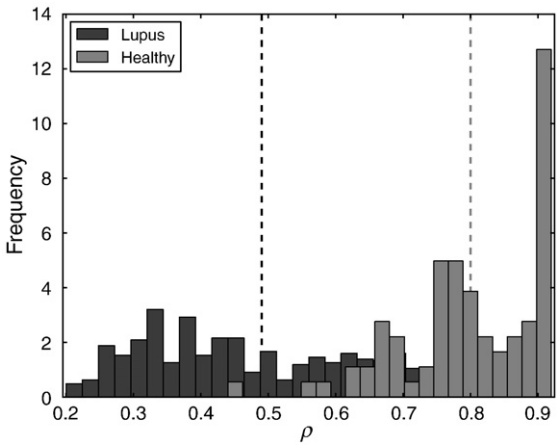

Fig. 8. Normed histograms of results from comparison of 600 Lupus samples and 85 Healthy to the Healthy reference set. For each sample, the metric ρ was determined and histograms were constructed from the results. Average (median) values of ρ are indicated with dashed lines.

are possible from the underlying constituent components, individual protein concentrations (amplitudes) and melting temperatures (peak centers) were varied systematically. Random perturbations were made by altering peak amplitudes and abscissa locations of four Gaussian curves which are sufficient for a first approximation to typical plasma thermograms. In total, 600 thermograms were generated that fell into either of two categories defined by the 90% quantiles of the established Healthy and Lupus reference sets.

The first set of simulated thermograms consisted of 300 curves generated by randomly selecting combinations of individual peak amplitudes so that the composite curve fell entirely within the 90% quantile range for the established Healthy reference thermograms. These curves are collectively referred to as 'simulated Healthy' since they were designed intentionally to be indistinguishable from Healthy reference curves. The second set of simulated thermograms consisted of 300 curves falling outside the 90% quantile range. These curves are collectively referred to as or 'simulated Diseased.' Representative thermograms from both sets of simulated curves are shown in Fig. 7 where differences in shape are clearly observable.

Each simulated thermogram was compared to the Healthy reference thermogram using the metrics described above. Average values of $P=0.899$ and $r=0.999$ were obtained for the simulated Healthy set and average values of $P=0.434$ and $r=0.799$ were obtained for the simulated Diseased set, resulting in similarity values of $\rho=0.948$ and $\rho=0.589$, respectively (see Table 1). With these results as a guide, a critical range of ρ values can be determined to calibrate the metric and direct thermogram classification processes. From these simulations, a low value of $\rho<0.6$ would indicate a test curve is at least as different from the Healthy reference curve as the average 'simulated Diseased' thermogram. Conversely, a high value of $\rho>0.8$ would indicate that a test curve is at least as similar to the Healthy reference thermogram as the average 'simulated Healthy' thermogram.

Due to the simplified, ideal nature of these simulations, where variations in overall thermogram shape are due only to changes in concentration of a few of the major proteins, correlations of reference and test thermograms are not impacted drastically. A more sophisticated perturbation model that simultaneously considers factors like interactions with ligands that also act to shift locations of the known protein peaks along the abscissa, in addition to individual peak shape would result in a broader range of ρ values.

3.3. Real case — healthy versus lupus

Similarity measurements were computed for large sets of both Healthy and Lupus thermograms. A collection of 85 Healthy thermograms not used to construct the Healthy reference set, and 600 thermograms from patients clinically diagnosed with lupus were compared to the Healthy reference set. For each sample, the metric P and correlation coefficient r were determined and values of ρ were calculated. Empirical optimization resulted in parameter values of $w=1.5$ and temperature range of 55–85°C. Histograms of resulting ρ values for both the Healthy and Lupus curves are shown in Fig. 8 and show that the composite ρ metric captures average differences in curve similarity for each type of sample. A relatively low average value of $\rho=0.487$ resulted for the Lupus samples that contrasted with the relatively high average value of $\rho=0.800$ calculated for the Healthy samples. Results are shown in Table 2.

3.4. Application — disease classification

When developing statistical methods for characterizing and classifying DSC thermograms, the ultimate goal is a quantitative (unsupervised) classification of unknown test thermograms into pre-characterized disease categories. To demonstrate this application, we classified a series of test curves into one of two disease categories (Healthy or Lupus) using standard classification techniques and the similarity index ρ described above.

In addition to the Healthy reference set of 100 Healthy thermograms characterized above, a collection of 300 Lupus thermograms were similarly characterized as a Lupus reference set. An additional 85

Table 1
Values of P, r and ρ computed by comparison of simulated thermograms to the Healthy reference set.

Thermogram type	P	r	ρ
Simulated Healthy	0.899	0.999	0.948
Simulated Diseased	0.434	0.799	0.589

Table 2
Average values of P, r and ρ computed by comparison of independent Healthy thermograms and thermograms obtained from confirmed lupus patients with the Healthy reference set.

Disease state	P_{avg}	r_{avg}	ρ_{avg}
Healthy	0.755	0.975	0.800
Lupus	0.425	0.759	0.487

Healthy thermograms and 300 Lupus thermograms not included in construction of the reference sets were subsequently used as test thermograms. Using the same temperature range and weighting parameter w above, each test curve $X_{test}(T)$ was compared to both Healthy and Lupus reference templates. Similarity to each template was measured using the metric ρ, and resulted in two values, $\rho_1(X_{test})$ and $\rho_2(X_{test})$, for each curve. The following standard decision rule was used to classify each test curve as either Healthy or Lupus:

Rule 1. If $\rho_1(X_{test}) > \rho_2(X_{test})$, classify X_{test} as 'Healthy', otherwise classify X_{test} as 'Lupus.'

That is, classify a test curve based on a comparison to whichever template is determined to be the most similar using the ρ metric. Applying this decision rule correctly classified 70 of the 85 (82.4%) known Healthy test curves and 265 of the 300 (88.3%) known Lupus test curves, which is a considerable success for such a simple metric. In fact, these values fall into a range that is entirely consistent with current diagnostic standards for antinuclear antibody testing for systemic lupus [11].

Development of a robust classification tool requires optimization of various aspects of the analytical process, including the pertinent criteria for the decision/classification rule. A more complete study that incorporates the statistical characterization and similarity methods described here, and investigates the utility of more advanced multivariate classification methods is currently underway and will be reported in future work. Nevertheless, the degree of success achieved with these simple methods is a testament to the rich information provided by thermal analysis of plasma samples and the utility of DSC as a potential medical research and clinical tool.

4. Discussion

The utility of analyzing and classifying DSC plasma thermograms lies in their categorization to intrinsic shape and pattern, with the primary objective being application to the diagnostic setting. Feasibility studies using DSC as a broadly applicable diagnostic platform have shown great promise and indicate that this method could have substantial clinical utility. However, ongoing data collection and analysis are required to firmly establish the clinical efficacy of this approach, as well as diagnostic specificity and sensitivity metrics.

The methods presented here for distinguishing plasma thermograms of healthy patients from those diagnosed with systemic lupus can, in principle, be applied to any disease condition that alters the plasma proteome in a manner detectable by DSC. Interactions within the plasma proteome that give rise to nontrivial changes in the melting temperatures of plasma proteins and result in DSC thermograms that are significantly different from the Normal baseline profile can be monitored and quantified using these techniques. Disease states that produce more subtle changes in the proteome profile that are still detectable by DSC but may fall below the threshold of sensitivity of these methods may require more sophisticated chemometric or statistical analysis in order to obtain comparable levels of accuracy and sensitivity. Nevertheless, these simpler methods are sufficient for detection and classification of major shifts and interactions in the plasma proteome produced by many disease states. Thus far we have investigated plasma thermograms from 25 different diseases, including a variety of cancers, different autoimmune diseases and neurological diseases. All diseases can be distinguished with statistical significance from normal controls, and each disease can be distinguished from all other diseases with a range of confidence levels.

Due to the highly complex nature of biological mixtures, few attempts have been made to address their composite melting behaviors. The ability to classify thermograms into distinct disease categories is taken as an indication that a common feature underlies the observed alterations in plasma thermogram profiles. Since a comprehensive physical model is currently lacking to support the diagnostic application, statistical methods have been developed to analyze and categorize thermograms based on empirical sets of reference thermograms.

DSC thermograms are sensitive to temperature dependent molecular transitions and/or chemical interactions between molecules that occur in solution. Such intra- and inter-molecular chemical binding events result in microscopic production or loss of heat that the DSC instrument collectively detects. Evidence suggests that many diseases may be characterized by interactions of small circulating ligands with plasma proteins present in high concentrations (e.g. Human Serum Albumin, IgG, Transferrin, etc) [1,2,12,13]. Thus it follows that many of these binding events may be responsible for variations in the shapes of observed plasma thermograms. By binding to proteins present in the sample, such ligands can act to shift the melting transition of one (or more) of the major protein peaks along the temperature axis. However, specific culprits responsible for observed shifts in certain diseases are not currently known, although studies to identify them are underway.

4.1. Physical models

Sophisticated theoretical models have been developed to help understand the melting process for homogeneous solutions of proteins or well defined ligand/protein mixtures [13–17]. In these models, closed-form equations describe reacting systems that contain both inter- and intra-molecular interactions (binding and melting, respectively). Although such physical models can provide powerful insights into the molecular constituents and interactions responsible for thermogram shifts, the idealized systems for which they have been developed do not take into account the possibility of multiplex reactions between elements in a biological sample (i.e. the interactome).

Plasma contains thousands of proteins ranging in concentration over nine orders of magnitude [18,19]. Due to the tremendous complexity of the system, formulation of a complete physical model considering all possible interactions is yet to be developed. Methods based on equilibrium and non-equilibrium thermodynamics [20] are currently being developed, which will lead to interpretations of plasma thermograms via functional deconvolution based on the effects of physical interactions between various reacting components. A sufficiently detailed model should allow more quantitative assessments of protein levels that could in turn serve as a guide for detecting disease biomarkers through their interactions with major plasma proteins.

4.2. Alternative methods

Computational techniques are commonly employed for complex signal analysis. Many well-known analytical methods exist to identify and compare patterns and shapes of various curve forms. Such methods have been successfully applied to diverse problems in fields ranging from biostatistics and biochemistry to atmospheric physics, seismology, materials science, computer science, and optics [21]. Although alternative analytical methods with similar capabilities could be employed, underlying assumptions and domain limitations (inherent or otherwise) often render these methods inappropriate or unnecessary for the specific analysis or classification of DSC thermograms.

A number of detailed computational procedures, including learning algorithms (such as neural networks or genetic algorithms), can be applied to the problem of curve pattern recognition [21–26]. Likewise, statistical optimization techniques can address a wide range of pattern analysis tasks including density estimation, clustering, feature selection and classification, error estimation or dimensional reduction. Non-parametric curve matching methods utilize various distance functions (like Taxicab or Hausdorff distances) or probabilistic measures of similarity to classify curves and shapes of curves [27,28]. Model-based curve matching techniques employ functional data analysis methods and employ numerical optimization routines to parameterize functional models to compare and classify curves and shapes [29,30]. Classical

fitting methods like the Kolmogorov–Smirnov, Cramer–von Mises, Chi-square, or Anderson–Darling tests are also often employed to quantify differences between data sets [31–33].

Many of these methods have been adapted to analyze DSC thermograms, but they are often either unnecessarily complex for the desired objectives, or involve assumptions regarding the data that we do not wish to invoke for an initial application. Because distributions of plasma thermogram data are not specifically known at this stage, parametric methods are not generally applicable. Moreover, many fitting methods are extremely sensitive to small differences in the observed data and so are not robust with respect to signal noise. Plasma thermogram reference sets have large variations that are unavoidable and fundamental to the instrumental system or system of measurement. Consideration of such variations is essential for analytical tasks to make accurate determinations of statistical significance and for establishing clinical utility.

Sophisticated multivariate statistical methods developed for chemometric analysis, such as principal component analysis, partial least-squares regression and linear discriminant analysis, are frequently employed for analysis and classification of data of the type presented here [34,35]. While these powerful techniques can be effective tools for classifying thermogram data, these methods are often strictly numerical and lack a clear physical interpretation. Although chemometric methods are extensively applied in similar settings, applications to DSC data is not common. Implementation of chemometric techniques to classify thermogram data is currently in development.

5. Conclusions

DSC thermograms present a powerful tool for identifying perturbations in plasma protein composition. These perturbations are reflected in differences (either subtle or dramatic) in thermogram patterns. Calorimetry provides a novel approach to diagnostic detection based on concentration, mass and thermodynamic stability that complements other commonly employed analytical techniques such as gel electrophoresis and mass spectrometry, which characterize plasma components based on size (mass), shape and charge. Mass spectrometry is sensitive to both size and charge while protein electrophoretic migration can be sensitive to size, shape and charge. In contrast, thermograms are derived from measurements of the heat generated or absorbed by the composite plasma solution as a function of temperature. It is this orthogonal viewpoint using DSC that provides a unique window into the plasma proteome.

The statistical methods presented here have been developed for general descriptive analysis and classification of thermogram data according to overall curve shape. An advantage of these methods is that they are easy to understand and implement without the need of more complex (or expensive) statistical software packages. Although ultimately applied to the problem of classifying DSC thermograms according to disease state, the characterization of reference sets and statistical comparisons produced by the metrics provide insight into the composition of plasma samples. This insight can help guide the development of more physical analytical techniques and thereby pave the way to accurate and sensitive diagnostic methods using the DSC-based plasma assay.

Acknowledgements

The authors are grateful to Dr. Anton Hopfinger and Dr. Alex Chagovetz for helpful advice, guidance and critical review. This work was supported in part by the National Science Foundation under the grant numbers IP-0912660 & IIP-1026824.

References

[1] N.C. Garbett, J.J. Miller, A.B. Jenson, J.B. Chaires, Calorimetry outside the box a new window into the plasma proteome, Biophysical Journal 94 (2008) 1377–1383.
[2] N.C. Garbett, J.J. Miller, A.B. Jenson, J.B. Chaires, Calorimetric analysis of the plasma proteome, Seminars in Nephrology 27 (2007) 621–626.
[3] N.C. Garbett, J.J. Miller, A.B. Jenson, D.M. Miller, J.B. Chaires, Interrogation of the plasma proteome with differential scanning calorimetry, Clinical Chemistry 53 (2007) 2012–2014.
[4] N.C. Garbett, C.S. Mekmaysy, C.W. Helm, A.B. Jenson, J.B. Chaires, Differential scanning calorimetry of blood plasma for clinical diagnosis and monitoring, Experimental and Molecular Pathology 86 (2009) 186–191.
[5] H. Wang, S.G. Clouthier, V. Galchev, D.E. Misek, U. Duffner, C.K. Min, R. Zhao, J. Tra, G.S. Omenn, J.L. Ferrara, S.M. Hanash, Intact-protein-based high-resolution three-dimensional quantitative analysis system for proteome profiling of biological fluids, Molecular & Cellular Proteomics 4 (2005) 618–625.
[6] S. Hanash, Disease proteomics, Nature 422 (2003) 226–232.
[7] E.P. Diamandis, Mass spectrometry as a diagnostic and a cancer biomarker discovery tool, Molecular & Cellular Proteomics 3 (2004) 367–378.
[8] J.D. Wulfkuhle, L.A. Liotta, E.F. Petricoin, Proteomic applications for the early detection of cancer, Nature Reviews. Cancer 3 (2003) 267–275.
[9] Z. Xiao, D. Prieto, T.P. Conrads, T.D. Veenstra, H.J. Issaq, Proteomic patterns their potential for disease diagnosis, Molecular and Cellular Endocrinology 230 (2005) 95–106.
[10] W.J. Krzanowski, Principles of Multivariate Analysis, Oxford University Press, New York, 1988.
[11] D.E. Habash-Bseiso, S.H. Yale, I. Glurich, J.W. Goldberg, Serologic testing in connective tissue diseases, Clinical Medicine & Research 3 (2005) 190–193.
[12] M. Zhou, D.A. Lucas, K.C. Chan, H.J. Issaq, E.F. Petricoin, L.A. Liotta, T.D. Veenstra, T.P. Conrads, An investigation into the human serum interactome, Electrophoresis 25 (2004) 1289–1298.
[13] R. Galletto, M.J. Jezewska, W. Bujalowski, Kinetics of allosteric conformational transition of a macromolecule prior to ligand binding analysis of stopped-flow kinetic experiments, Cell Biochemistry and Biophysics 42 (2005) 121–144.
[14] J.F. Brandts, L.N. Lin, Study of strong to ultratight protein interactions using scanning calorimetry, Biochemistry 29 (1990) 6927–6940.
[15] T.T. Waldron, K.P. Murphy, Stabilization of proteins by ligand binding application to drug screening and determination of unfolding energetics, Biochemistry 42 (2003) 5058–5064.
[16] M.S. Celej, S.A. Dassie, M. Gonzalez, M.L. Bianconi, G.D. Fidelio, Differential scanning calorimetry as a tool to estimate binding parameters in multiligand binding proteins, Analytical Biochemistry 350 (2006) 277–284.
[17] S.J. Gill, Thermodynamics of ligand binding to proteins, Pure and Applied Chemistry 61 (1989) 1009–1020.
[18] N.L. Anderson, N.G. Anderson, The human plasma proteome history, character, and diagnostic prospects, Molecular & Cellular Proteomics 1 (2002) 1845–1867.
[19] J.N. Adkins, S.M. Varnum, K.J. Auberry, R.J. Moore, N.H. Angell, R.D. Smith, D.L. Springer, J.G. Pounds, Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry, Molecular & Cellular Proteomics 1 (2002) 1947–1955.
[20] M.T. Horne, D.J. Fish, A.S. Benight, Statistical thermodynamics and kinetics of dna multiplex hybridization reactions, Biophysical Journal 91 (2006) 4133–4153.
[21] A.K. Jain, R.P.W. Duin, J. Mao, Statistical pattern recognition a review, J IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (2000) 4–37.
[22] J.R. Parker, Scientific curve classification by combining simple algorithms, in: Proceedings of the First IEEE International Conference on Cognitive Informatics (ICCI'02), 2002.
[23] C.G. Small, H. Le, The statistical anlaysis of dynamic curves and sections, Pattern Recognition 35 (2002) 1597–1609.
[24] S. Manay, D. Cremers, A. Yezzi, S. Soatto, One-shot integral invariant shape priors for variational segmentation, Lecture Notes in Computer Science, Energy Minimization Methods in Computer Vision and Pattern Recognition, volume 3757, Springer-Verlag, Berlin Heidelberg, 2005, pp. 414–426.
[25] A. Webb, Statistical Pattern Recognition, 2nd Edition, Wiley, 2002 Reprint September 2004.
[26] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, Wiley, 2001.
[27] R. Moeckel, B. Murray, Measuring the distance between time series, Physica D 102 (1997) 187–194.
[28] A. Bowman, S. Young, Graphical comparison of nonparametric curves, Applied Statistics 45 (1996) 83–98.
[29] Y.M. Maldonado, J.G. Staniswalis, L.N. Rwin, D. Byers, A similarity analysis of curves, The Canadian Journal of Statistics 30 (Sep 2002) 373–381.
[30] H. Motulsky, A. Christopoulos, Fitting models to biological data using linear and nonlinear regression, A practical guide to curve fitting, GraphPad Software Inc, San Diego CA, 2003.
[31] M.A. Stephens, Use of the kolmogorov-smirnov, cramer-von mises and related statistics without extensive tables, Journal of the Royal Statistical Society: Series B: Methodological 32 (1970) 115–122.
[32] I.T. Young, Proof without prejudice: use of the Kolmogorov–Smirnov test for the analysis of histograms from flow systems and other sources, The Journal of Histochemistry and Cytochemistry 25 (1977) 935–941.
[33] Y. Arzhaeva, D.M.J. Tax, B. Van Ginneken, Dissimilarity-based classification in the absence of local ground truth: application to the diagnostic interpretation of chest radiographs, Pattern Recognition 42 (2009) 1768–1776.
[34] K. Varmuza, P. Filzmoser, Multivariate Statistical Analysis in Chemometrics, CRC Press, Boca Raton, Fl, 2009.
[35] R. Kramer, Chemometric Techniques for Quantitative Analysis, Marcel Dekker, NY, 1998.