Biochimica et Biophysica Acta xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagen

Differential scanning calorimetry as a complementary diagnostic tool for the evaluation of biological samples

Nichola C. Garbett [a,*], Guy N. Brock [b,**]

[a] James Graham Brown Cancer Center, Department of Medicine, University of Louisville, Louisville, KY, USA
[b] Department of Bioinformatics and Biostatistics, School of Public Health and Information Sciences, University of Louisville, Louisville, KY, USA

ARTICLE INFO

Article history:
Received 28 August 2015
Received in revised form 7 October 2015
Accepted 8 October 2015
Available online xxxx

Keywords:
Diagnostic classification
Differential scanning calorimetry
Plasma proteome
Systemic lupus erythematosus
Thermogram
Penalized classification methods

ABSTRACT

Background: Differential scanning calorimetry (DSC) is a tool for measuring the thermal stability profiles of complex molecular interactions in biological fluids. DSC profiles (thermograms) of biofluids provide specific signatures which are being utilized as a new diagnostic approach for characterizing disease but the development of these approaches is still in its infancy.
Methods: This article evaluates several approaches for the analysis of thermograms which could increase the utility of DSC for clinical application. Thermograms were analyzed using localized thermogram features and principal components (PCs). The performance of these methods was evaluated alongside six models for the classification of a data set comprised of 300 systemic lupus erythematosus (SLE) patients and 300 control subjects obtained from the Lupus Family Registry and Repository (LFRR).
Results: Classification performance was substantially higher using the penalized algorithms relative to localized features/PCs alone. The models were grouped into two sets, the first having smoother solution vectors but lower classification accuracies than the second with seemingly noisier solution vectors.
Conclusions: Coupling thermogram technology with modern classification algorithms provides a powerful diagnostic approach for analysis of biological samples. The solution vectors from the models may reflect important information from the thermogram profiles for discriminating between clinical groups.
General significance: DSC thermograms show sensitivity to changes in the bulk plasma proteome that correlate with clinical status. To move this technology towards clinical application the development of new approaches is needed to extract discriminatory parameters from DSC profiles for the comparison and diagnostic classification of patients.

1. Introduction

Differential scanning calorimetry (DSC) is an established biophysical technique that monitors heat capacity changes associated with the thermal denaturation of biomolecules. DSC is the only technique that directly measures the thermodynamics of intra- and inter-molecular interactions stabilizing biological systems. It is a very powerful technique that has been applied to the characterization of biomolecules in a wide range of applications. A major focus has been on the rigorous determination of thermodynamic driving forces governing protein stability, folding and binding interactions as well as more qualitative thermal stability studies for drug development and biopharmaceutical formulations (e.g. [1,2]). More recently, DSC has been applied in a new direction to the analysis of complex biological samples, such as blood plasma or other biofluids [3]. A growing number of studies suggest that differences in the thermodynamic properties of biofluid proteomes can be used to differentiate clinical samples based on health status [4–33]. It might seem initially surprising that DSC would display such sensitivity to changes in the bulk, high abundance proteome as a result of disease processes. In fact, our lab has reported that DSC changes are not related to differences in the concentration of major blood proteins in disease plasma samples. However, given the high sensitivity of DSC to modulation of biomolecule stabilization through changes in intra- and inter-molecular interactions, it is plausible that DSC biofluid profiles could

Abbreviations: AUC, area under the curve; C_p^{ex}, excess specific heat capacity; DSC, differential scanning calorimetry; GLM, generalized linear model; LASSO, least absolute shrinkage and selection operator; LDA, linear discriminant analysis; LDA-LASSO, lasso constrained LDA model; LFRR, Lupus Family Registry and Repository; LR, logistic regression; LR-LASSO, lasso constrained LR model; LR-ENET, elastic net constrained LR model; PC, principal component; PLS, partial least squares; PLS-DA, PLS discriminant analysis; ROC, receiver operating characteristic; SCAD, smoothly clipped absolute deviation; SLE, systemic lupus erythematosus; SPLS-DA, sparse PLS-DA; SVM, support vector machines; SVM-ESCAD, elastic SCAD penalized SVM model; SVM-SCAD, SCAD penalized SVM model; T_{FM}, first moment temperature; T_{max}, temperature of the peak maximum.

* Correspondence to: N.C. Garbett, James Graham Brown Cancer Center, Department of Medicine, University of Louisville, Clinical and Translational Research Building, Room 206, 505 South Hancock Street, Louisville, KY 40202, USA
** Correspondence to: G.N. Brock, Department of Bioinformatics and Biostatistics, School of Public Health and Information Sciences, University of Louisville, Room 137, 485 East Gray Street, Louisville, KY 40202, USA.
E-mail addresses: nichola.garbett@louisville.edu (N.C. Garbett), guy.brock@louisville.edu (G.N. Brock).

http://dx.doi.org/10.1016/j.bbagen.2015.10.004
0304-4165/

reflect the modified thermal stability of major proteome components resulting from covalent modifications or binding interactions involving disease biomarkers. Although the biological mechanisms of DSC changes have not yet been reported, compelling evidence suggests the potential utility of DSC for clinical diagnostics in multiple disease settings. To realize the possibility of the clinical application of DSC, the challenge lies in demonstrating the reliable technical performance of the assay and in developing approaches to extract diagnostic information from the complex DSC profiles that can be easily applied to yield straightforward clinical metrics.

To facilitate interpretation of DSC data of clinical samples, a number of studies [9,10,12,13,16,17,19,23,24,26] have reported the calculation of metrics that provide a read-out of specific localized features of DSC profiles (e.g. heat capacity and temperature maxima of profiles). These features have been useful in discerning trends in clinical groups and calculating the statistical significance of these differences. To utilize information from the entire thermogram for diagnostic classification a number of global analysis methods have been developed. One approach used a non-parametric method to determine differences between DSC profiles based on the distance between a test profile and averaged profiles for each class [15]. The distance was defined as the geometric average of the correlation between DSC profiles (i.e. similarity in shapes) and Euclidean distance. This approach was used to analyze DSC profiles of healthy controls and lupus patients and achieved 82% correct classification of healthy profiles and 88% for lupus. The method was developed for the analysis of plasma DSC profiles for the lupus/control data set but is generally applicable to other data sets of biofluid DSC profiles in any disease setting where any test profile can be compared to a well-defined reference group. Other groups have applied this approach for the analysis of DSC data in different settings, for example, for the classification of colorectal cancer based on DSC profiles [24]. Another method for the analysis of plasma profiles employed a parametric statistical model developed for the classification of cervical cancer versus healthy controls [22]. Here, DSC profiles were reduced in complexity by restricting the temperature range to that encompassing the major heat capacity signal (50–76°C) and averaging over 1°C temperature increments. Profiles were then subjected to a logarithmic transformation and fit to a linear regression model. This method performed extremely well for the healthy/cervical cancer data set with a mean classification rate of 97%. As the model used in this approach was developed from this specific data set, the utility of this method requires the evaluation of this approach with other data sets. Another useful method was based on deconvoluting the DSC profile into several component curves, each with a defined height, center and width which were used in a multiparametric analysis for the classification of healthy controls and gastric adenocarcinoma patients [27]. The construction of polygonal plots from these three parameters for each of the component curves provided a useful graphical tool to distinguish patient groups. Also, similar to some earlier reports, the area and first moment, or average, temperature of DSC profiles were found to display differences between the controls and gastric adenocarcinoma patients.

The approaches discussed above demonstrate the evolution in the development of analytical tools for the characterization of DSC biofluid profile features associated with various clinical conditions. More work is needed to develop and validate reliable analytical approaches to provide a rapid, easily interpretable diagnostic result that can be readily employed in the clinical setting. The focus of this report is the application of new analysis approaches to DSC data for the purpose of diagnostic classification.

2. Materials and methods

2.1. Plasma samples

De-identified plasma samples and patient data were obtained from the Lupus Family Registry and Repository (LFRR) [34]. Plasma samples for 300 patients meeting the revised criteria of the American College of Rheumatology for SLE [35] and 300 healthy controls matched demographically by sex, ethnicity and age were received and kept at −80°C until thawed for DSC analysis.

2.2. Collection of DSC thermograms

DSC samples were prepared and analyzed according to our previously published procedure [10] which includes a detailed account of our experimental procedures. Data were collected using an automated VP–Capillary DSC system (MicroCal, LLC, Northampton, MA, now a division of Malvern Instruments Inc.). Electrical calibration of the differential power signal and temperature calibration using hydrocarbon temperature standards were performed as part of the manufacturer periodic instrument maintenance. Interim instrument performance was assessed using biological standards lysozyme and RNaseA. Samples and dialysate were loaded into 96 well plates thermostated at 5°C within the instrument autosampler until analysis. Thermograms were recorded from 20°C to 110°C at a scan rate of 1°C/min with a pre-scan thermostat of 15 min, mid feedback mode and a filtering period of 2 s. Duplicate thermograms were obtained for each plasma sample. DSC data were analyzed using Origin 7 (OriginLab Corporation, Northampton, MA). Raw DSC data were corrected for the instrumental baseline by subtraction of a suitable buffer reference scan. Thermograms were normalized for the total protein concentration and corrected for non-zero baselines by application of a linear baseline fit. Final thermograms were plotted as excess specific heat capacity (J/K · g) versus temperature (°C).

2.3. Summary metrics of DSC thermograms

Thermograms are frequently characterized by metrics summarizing the shape and prominent features of the thermograms [9]. These include: (1) the total area under the thermogram (typically from 45 to 90 °C); (2) the maximum excess specific heat capacity at various peaks (e.g. Peak 1 height, Peak 2 height); (3) the overall maximum peak height; (4) the width of the primary thermogram peak at half height; (5) the temperature of the peak maximum (T_{max}); (6) the ratio of the peak heights (e.g., (Peak 1 height)/(Peak 2 height)); and (7) the "mean" or first moment temperature of the thermogram, T_{FM}, where

$$T_{FM} = \frac{\int_{45}^{90}\left(TC_p^{ex}\right)dT}{\int_{45}^{90}C_p^{ex}dT}$$

and C_p^{ex} represents the excess specific heat capacity at a given temperature. These summary metrics can be used in lieu of the original thermogram values for classifying disease status based on any of the classification models described below. While the calculated metrics are useful for characterizing certain aspects of the thermograms, they are not necessarily informative for differences between patient classes.

Principal components (PCs) are another common technique for summarizing the information in a data matrix in a concise manner. PCs are not intended as a classification technique per se, but are commonly used as a dimension reducing tool prior to building a classification or regression model. PCs are the set of orthogonal vectors or factors such that the first vector is the direction which explains the most variation in the data, the second vector is the direction which explains the second greatest percentage of variation in the data that is orthogonal to the first, and so on. The solution can be obtained from the eigenvalue decomposition of the covariance matrix of the data, where the principal components directions are the eigenvectors of the covariance matrix and the variance of the principal components are proportional to the eigenvalues of the covariance matrix. Typically, only the number of components needed to explain 90–95% of the total variation in the data are retained (e.g., as determined by the "elbow" in a scree

Please cite this article as: N.C. Garbett, G.N. Brock, Differential scanning calorimetry as a complementary diagnostic tool for the evaluation of biological samples, Biochim. Biophys. Acta (2015), http://dx.doi.org/10.1016/j.bbagen.2015.10.004