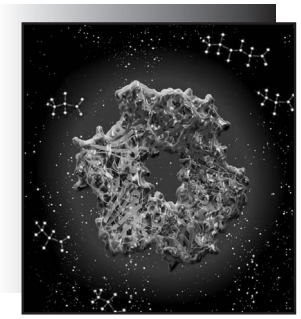

Quantitative Analysis of Proteomics Using Data Mining

An Automated System for Constructing Assays Quickly and Precisely

BY CHIA-YU YEN, STEVE M. HELMKE, KRZYSZTOF J. CIOS, M. BENJAMIN PERRYMAN, AND MARK W. DUNCAN

An important problem in biology is to determine the concentration of a protein of interest contained in a complex protein mixture. The amount of a particular protein may be useful for disease diagnosis or determining patient prognosis. Proteomics provides a systematic way to solve this problem [1], [2].

Two closely related protein isoforms of myosin heavy chain (MyHC) designated α and β are expressed in the human heart. Changes in α- and β-MyHC mRNA concentrations have been implicated in heart failure [3], [4] and a positive response to therapy [5], but changes in mRNA concentrations do not always correlate with changes in protein concentrations [6], [7]. Protein concentrations are more relevant because protein carries out the actual function. It is known that contractile velocity is affected by the relative amount of these two proteins. The failing human heart has less α-MyHC [8]. Thus, determining the relative amount of these proteins is important for determining patient prognosis. Unfortunately, the α- and β-MyHC proteins cannot be easily distinguished using conventional electrophoretic or immunological approaches. An assay using matrix-assisted laser desorption/ionization (MALDI) time-of-flight (TOF) mass spectrometry (MS) to measure protein concentrations by quantifying specific tryptic peptides from these closely related protein isoforms has recently been developed [9].

The proposed quantification method requires an internal standard peptide. A stable isotope-labeled version of a peptide is the simplest internal standard peptide to create. Stable isotope-labeled standards are commonplace in the quantitative mass spectrometry analyses of small molecules [10]. However, the cost of a stable isotope-labeled peptide is very high. Therefore, we adopted a different method, which requires an unlabeled peptide for quantifications of two proteins. This quantification process is as follows.

- ➤ Find protein isoforms that change with heart failure.
- ➤ Find tryptic peptides that distinguish each protein isoform and have strong MS signals and designate these peptides quantification peptides.
- ➤ Design an internal standard peptide derived from the quantification peptides.
- ➤ Synthesize quantification peptides and the internal standard peptide.
- ➤ Mix known amounts of these synthetic peptides for MS.
- ➤ Create standard curves from these MS data signals.
- ➤ Add a known amount of the internal standard peptide to tryptic digests of samples from archived patient heart tissue.
- ➤ Collect MS data signals of quantification peptides and the internal standard peptide.
- ➤ Use the collected signals to measure the amounts of the two protein isoforms.

In absence of automated systems, the quantification process and the tasks of finding quantification peptides, designing the internal standard peptide, processing the MS data, calculating standard curves, and quantifying sample MS data are very labor intensive.

Our goal was to develop an automated system that would ensure a robust peptide quantification process, which would permit researchers to quantify desired proteins faster and with greater reliability. Because of the uniqueness of the data used, we employed both biochemists' expertise and data mining methods in this work [11], [12]. The system includes two main system components: one for the discovery of two quantification peptides and the internal standard peptide and the other for protein quantification in patient samples. If the required input data are available, each subsystem can be run separately. The developed system can be applied to similar problems because our design is flexible, allowing for easy adaptation.

Mining Information for the Quantification and Internal Standard Peptide

Careful selection of the quantification peptide and design of the internal standard peptide are crucial for accurate quantitative MALDI-TOF MS data. This section describes the subsystem that can help biologists identify quantification peptides and the internal standard peptide (IS) by using biological background knowledge. The subsystem requires knowledge about amino acid families to determine the optimal conservative amino acid substitution. With this information available, the subsystem finds possible quantification peptide pairs and designs an internal standard peptide for each pair.

Finding the Quantification Peptide

For a given protein, a quantification peptide must be selected specifically and uniquely for that protein in the context in which it will be measured. For example, a highly conserved

 0739-5175/05/$20.00©2005IEEE

protein such as human cardiac α-MyHC would have quantification peptides shared with other species; however, if only human samples were analyzed, then the quantification peptide would have to discriminate human cardiac α-MyHC from other human cardiac myosin isoforms.

There are three rules that we use for finding the pair of quantification peptides [9].

a) The peptides differ by a single conservative amino acid substitution.
b) The tryptic cleavage site environments of these peptides are identical.
c) The peptides have strong MS signals.

Figure 1 illustrates the first two rules [9].

Required Data

To find quantification peptides, five input files are needed: two protein sequence files and three MALDI-TOF MS data files. Each protein sequence file contains the amino acid sequence of one of the protein isoforms. By assumption, the user already knows the target proteins and obtains the target protein sequences from protein databases such as Swiss-Prot or NCBInr. The three MALDI-TOF MS data files contain the MS data of the first isoform, of the second isoform, and a representative sample of the material to be analyzed (which contains a mixture of the isoforms). The MS data files of the isoforms can be from pure proteins or from preparations enriched in one of the isoforms and are used to determine the MS signal intensities of the candidate quantification peptides. The MS data file of the representative sample is used to find open regions for placement of the internal standard peptide.

Algorithm for Finding Quantification Peptides

In order to find the desired quantification peptides, an algorithm is designed based on the above described background knowledge.

If all the rules are applied at the same time, all peptide pairs need to be evaluated. Therefore, the algorithm uses the rules in the order specified above. The algorithm is designed to show not only the peptide pair with the strongest signals but also all pairs fulfilling other rules. Rule c) is used to sort the identified candidates by signal strength, which means it will not reduce the number of candidates. Thus, rule c) is used last. Substitution in rule a) means replacing one amino acid for another, implying that the length of the two peptides will be the same. As a result, rule a) not only determines the conservative amino acid's substitution but also compares the lengths of two peptides, meaning it can eliminate more peptide pairs than rule b). Therefore, we use rule a) as the first rule. In rule b), the cleavage site environment is defined as the two amino acids before the cleavage site and the two amino acids after the cleavage site.

The pseudocode of the above outlined algorithm is

```
INPUT S_α: Protein 1, S_β: Protein 2, MS_α: α MS, MS_β: β MS

P_α ← Apply in silico digestion to S_α
P_β ← Apply in silico digestion to S_β

For each peptide pair of p_α ∈ P_a AND p_β ∈ P_β
  If HammingDistance (p_α, p_β) = 1 AND
    differed AAs within the same AA family
    Then
      Candidates.insert(p_α, p_β)
  End-If
End-For

For each pair IN Candidates
  If the two AAs before and after the cleavage sites of both the peptides are different Then
      Remove this pair from Candidates
  End-If
End-For

For each pair IN Candidates
  p_α.intensity ← MS_α.getIntensity(p_α.mass)
  p_β.intensity ← MS_β.getIntensity(p_β.mass)
End-For

Sort Candidates by intensity p_α and p_β

Return Candidates
```

Protein sequence alignment algorithms such as BLAST [13] and the dynamic programming pairwise sequence alignment are currently available [14], [15]. These algorithms can calculate the similarity of a pair of sequences by using the accepted point mutation (PAM) matrix [16], and they can find a score to indicate the relationship between the sequences. We chose a straightforward approach that checks the length of the given pair of peptides and determines the relationship of the amino acids at the substitution site.

Designing the Internal Standard Peptide

The criteria for a peptide to be the internal standard peptide are as follows [9].

- The sequence of this peptide must be altered from that of the quantification peptide so that the mass of this peptide can be discriminated by MALDI-TOF MS while maintaining the chemistry of the original quantification peptide.
- The substitution should not change the charge or hydrophobicity of the peptide, as this would change the recovery of the peptide or the ability of the peptide to cocrystallize with the matrix or the ability to ionize, and therefore change, the production of its MALDI-TOF signal.
- The mass of the internal standard peptide should be in an open region of the spectrum in which the peptide signal can appear without

		m/z [M+H]$^+$ Mono-Isotopic
α-MyHC 726-741 Peptide:	YR|ILNPVAIPEGQFIDSR|KG	1,768.96
Internal Standard Peptide:	ILNPVAVPEGQFIDSR	1,754.94
β-MyHC 724-739 Peptide:	YR|ILNPAAIPEGQFIDSR|KG	1,740.93
	↑ Trypsin Site ↑ Trypsin Site	

Fig. 1. Myosin heavy chain quantification peptides (from (9)).

interference from other peptides. This open region must be near the quantification peptide since the peptide will have a mass close to that of the quantification peptide.

- ➤ If there are several potential quantification peptides, then the sample spectra can be inspected to find the quantification peptides that have the highest signal and also have nearby open regions for the internal standard peptide signal.

The most important of these rules is to maintain the chemical properties of the internal standard peptide as close as possible to the original peptide. We achieve this using a conservative amino acid substitution [9]. The other critical component is to find an open region of the spectrum where a synthetic peptide can be inserted without interference from the sample.

Algorithm for Designing Internal Standard Peptide

This algorithm uses the above stated rules to find the candidates for internal standard peptides and takes quantification peptides and a spectrum including quantification peptide peaks for open region determination as parameters.

First, the algorithm "guesses" where the internal standard peptide signal should be located. The reasonable prediction is the center between the two quantification peptides. Assume, that the given quantification peptide pair is (p_α, p_β). The mass difference between p_α and this predicted mass value is denoted as D. We know that the substitutions of any two amino acids within an amino acid family may not cover all found values of D. Because of this limitation, the algorithm finds all possible alternative distances by given amino acid families and uses the one closest to D, instead of the assigned guess.

Second, the algorithm checks the existence of an open region. In order not to consider noise present in the MS data we make the following assumptions. From observations (Figure 2), in any MS data file more than 90% of the peaks have intensities less than 10%, so we use 10% as the average noise, and a factor of 1.5 as possible boundary for all noise. Due to the presence of isotopes, a range wide-enough to insert an internal standard and its isotopes is inspected. If all peaks in this range are less than the noise boundary, it can be considered an open region. When the predicted internal standard can be inserted into the sample without interference, the algorithm goes to the next step; otherwise it stops.

In the last step of the algorithm we need to check for the conservative amino acid substitution. The algorithm substitutes amino acids from the site where p_α and p_β are different on both sides. There are two biological considerations behind this. The first one is that the remaining part of the internal standard peptide will be very similar to both quantification peptides, so it could be used to quantify either isoform. The other is that the end amino acids determine more of the peptide's chemistry. This means that the end substitution of a peptide will probably change chemical properties of a peptide more than substitutions at an internal site. The algorithm will only check for a possible substitution within given amino acid families. Once it finds a site that satisfies the conditions of mass value and the conservative substitution, a new candidate peptide sequence is created using the identified substitution. The pseudocode of the algorithm follows.

```
INPUT pα: quant pep 1, pβ: quant pep 2, MSp: Patient MS

  D ← |pα.mass−|pα.mass−pβ.mass|/2|
  D ← FindPossible AADifferential(D)
  MIS ← pαmass + D

If MIS not IsInOpenSpace (MIS) Then
  Return null
End-If

S ← Possible AASubstitutions (D)

For each amino acid AA IN pα
  If S.contains(AA) Then
      AA'← S.getSubstitution (AA)
      IS ← pα.substitute (AA,AA')
      Candidates.append(IS)
  End-If
End-For

Return Candidates
```

All the steps are based only on p_α. In order to find all possible internal standard peptides, p_α and p_β are exchanged and the process is repeated.

Quantification of Mass Spectrometry Data

In this section we describe a subsystem that quantifies MS data automatically. Before quantifying the experimental MS data, we

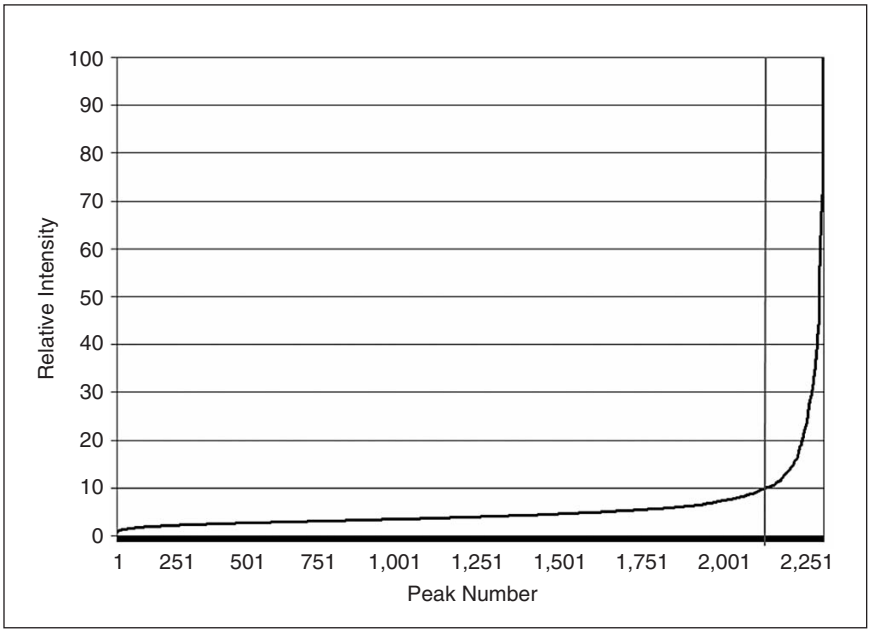

Fig. 2. MS data sorted by the relative intensity.

first obtain the necessary information in terms of the quantification peptides and the internal standard peptide. They can be determined by the subsystem described in the previous section.

The required information is the isotopes of the quantification peptides, the isotopes of the internal standard peptide, and the standard curves. The reason to consider the intensities, or amounts of isotopes, is that sometimes the intensities of isotopes are almost the same as, or greater than, the monoisotopic peak intensity. If isotopes are ignored, a large portion of the MS signal data may be discarded. Using this information, MS data can now be quantified. The data we use here were collected from patient heart tissues provided by Helmke et al.

Standard Curves

The standard curve is a very important factor in the quantification process. Without the standard curve, the user cannot calculate the quantification result from mass spectrometry data. The standard curve data are prepared from synthetic peptides, not from a patient sample.

The subsystem calculates two different standard curves: one is the ratio of quantification peptides, and the other is the ratio of the quantification peptide to the internal standard. The first can be used to find the relative quantification of two quantification peptides. The second is used to calculate the absolute amount of each quantification peptide.

In order to adapt different data sets, which might not be fit by linear regression, the nonlinear regression model is chosen:

$$y = f(x, \theta) + \varepsilon, \tag{1}$$

where θ represents unknown parameters. A nonlinear least-squares function is used to estimate the unknown parameters [17], [18]

$$S(\theta) = \sum_{i=1}^{n} [y_i - f(x_i, \theta)]^2. \tag{2}$$

The basic idea is to evaluate the residual sum of the squares function $S(\theta)$ iteratively. The set of unknown parameters which has the smallest evaluation value is chosen as the parameters of function f.

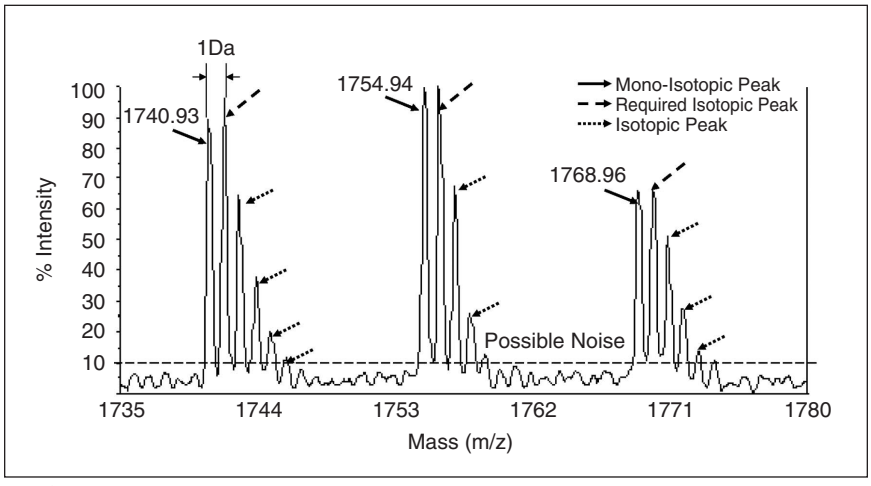

Fig. 3. An illustration of the strategy for choosing the desired peaks.

There are four regression types implemented: linear, quadratic, cubic, and fourth order. When the regression type is chosen, the subsystem will use it as function f in (1). The user can evaluate how well the chosen type fits the input data by fitting Goodness or the coefficient of determination R^2 [17], [18].

$$\text{Goodness} = R^2 = 1 - \frac{\sum_{i=1}^{n} [y_i - f(x_i)]^2}{\sum_{i=1}^{n} (y_i - y_{\text{mean}})^2} \tag{3}$$

Strategy for Choosing the Peaks

The data of each experiment include many peaks, but not all the peaks and their isotopes are useful. Hence, only monoisotopic peaks of quantification peptides and the internal standard peptide and their isotopic peaks, the number of those peaks defined by users, are required.

Some criteria have to be met when the subsystem collects the peak data. The first one is the error range. Because of experimental error, the measured mass values of the peaks are not exactly the same as the calculated mass values. Therefore, for each peak, an error range is set for peak determination. We chose 0.5 Da as the standard error range in this subsystem because the mass values of isotopic peaks are 1 Da higher than its previous peak.

The second criterion is the existence of the required peaks. If the user wants to use data from three isotopic peaks, the subsystem needs to sum up the intensities (peak heights) of the monoisotopic peak and all required isotopic peaks. In a situation where some of these peaks are not detected, the subsystem assumes that a quantification peptide or an internal standard peptide is not abundant enough to detect in the sample. The rationale is that a peptide should have monoisotopic and isotopic peaks detected in MS data. In other words, some consecutive peaks cannot be treated as the peptide peaks when these peaks lack either the monoisotopic peak or the required isotopic peaks. Therefore, the abundance of these peaks cannot be counted in the data calculation.

The third criterion is the comparison of the intensity of each detected peak with background noise. If any required peak of a quantification peptide is less than background noise, the subsystem treats this quantification peptide as "detected but too small" and will not use this value for calculation. The last criterion applies if there are two peaks within the same range. In this case, the subsystem takes only the highest one for calculation, as this is the true peak and the lower peak is a shoulder to the main peak. Figure 3 illustrates these considerations.

Quantification Data Calculation

Using the strategy outlined above, the required intensities of two quantification peptides and an internal standard peptide are known. Now, let H_α, H_β, and H_{IS} be the intensities of quantification peptides P_α, P_β, and internal

standard peptide P_{IS}. First, the signal ratio of P_α and P_β are calculated and denoted as

$$\%P_\alpha MS = \frac{H_\alpha}{H_\alpha + H_\beta}. \quad (4)$$

Let the standard curve of the ratio of the quantification peptides be f_R. The inverse estimation of f_R is used to determine the quantification peptide ratio of these two peptides and the result is

$$\%P_\alpha \ Peptide \ = f_R^{-1}(\%P_\alpha MS). \quad (5)$$

Next, the ratios of P_α/P_{IS} and P_β/P_{IS} must be found. With these ratios, the same as above, we use the inverse function of the standard curves f_P of P_α/P_{IS} and P_β/P_{IS} to estimate the amounts of P_α and P_β

$$\text{pmol } P_i = f_{P_i}^{-1}(H_i/H_{IS}),\ i = \{\alpha, \beta\}. \quad (6)$$

By providing the total amount W of P_α and P_β, the amount of peptide per microgram is calculated as

$$\text{pmol } P_i/\mu g = \frac{\text{pmol } P_i}{W},\ i = \{\alpha, \beta\}. \quad (7)$$

Error Estimation

Because the results of the quantification process are found by inverse functions of the standard curves, we estimate the error range using standard deviation. We also implemented the confidence interval to determine the inverse error range [17]. The advantage of the latter is that it is more precise but works only for linear regression.

Quantifying Human Heart MyHC Results

Changes in α- and β-MyHC mRNA were observed in human heart failure [3], [4]. Therefore, we developed an assay to measure α- and β-MyHC protein [9], and the experimental MS data from this work constitutes the input to our system. As part of this project we selected quantification peptides and designed an internal standard peptide (see Figure 1). These peptides, found by the human investigators, were used to validate the correctness of this automated system. In addition, manually calculated quantification results were used to confirm the output of our system.

Quantification Peptides and Internal Standard Peptide

The manually found quantification peptides and the internal standard peptide are as shown in Figure 1. Table 1 shows detailed information of found quantification pairs in descending order of MS signal intensity.

Although several candidate pairs were found by the system, only the first pair was used according to the previously described rules. Table 2 shows the internal standard peptide candidates for the peptide pair with the strongest corresponding MS signal.

In a comparison of the manually found data (Figure 1) with automatically found peptides (Tables 1 and 2), the first pair of quantification peptides in Table 1 shows that the quantification peptide found from β-MyHC (the lower protein peptide) is ILNPAAIPEGQFIDSR (this peptide is located in the protein from site 726 to 741 and its monoisotopic mass value is 1,740.93) and that the quantification peptide found from α-MyHC (the higher protein peptide) is ILNPVAIPEGQFIDSR (this peptide is located in the protein from site 724 to 739, and its monoisotopic mass value is 1,768.96).

The sequences and locations of these two peptides found automatically are the same as the peptides shown in Figure 1, validating the system.

As for the internal standard (IS) peptide, the first IS peptide in Table 2 is still the same as the given data because it matched the specified rules. The second IS peptide also matched all of the rules but the substitution site is too far from the difference site (Site 5) found in the two quantification peptides. Our proteomics experts thought that the second substitution site might also be appropriate because it matches all the

Table 1. Possible quantification peptide pairs.

Quantification peptide of β-MyHC			Quantification peptide of α-MyHC		
Sequence	**Mass**	**Site**	**Sequence**	**Mass**	**Site**
(YR)ILNPAAIPEGQFIDSR(K)	1,740.9281	724–739	(YR)ILNPVAIPEGQFIDSR(K)	1,768.9594	726–741
(K)ALQEAHQQALDDLQAEEDK(VN)	2,152.0146	998–1,016	(K)ALQEAHQQALDDLQVEEDK(VN)	2,180.0459	1,000–1,018
(VK)EDQVMQQNPPK(FD)	1,313.6151	73–83	(VK)EDQVLQQNPPK(FD)	1,295.6587	73–83
(SR)NEALR(VK)	602.3262	1,609–1,613	(SR)NEVLR(VK)	630.3575	1,611–1,615

Table 2. Possible internal standard peptides.

Quantification peptide of β-MyHC		Internal Standard			Quantificaiton peptide of α-MyHC	
Sequence	**Mass**	**Sequence**	**Mass**	**SubstitutionSite**	**Sequence**	**Mass**
ILNPAAIPEGQFIDSR	1,740.9281	ILNPVAVPEGQFIDSR	1,754.9438	7	ILNPVAIPEGQFIDSR	1,768.9594
ILNPAAIPEGQFIDSR	1,740.9281	ILNPVAIPEGQFVDSR	1,754.9438	13	ILNPVAIPEGQFIDSR	1,768.9594

Table 3. Quantification result of patient sample.

Patient 1 IS: 2.0 pmol								pmol/μg			pmol/μg	pmol/μg
Filename	β-MyHC	IS	α-MyHC	% α-MS	% α-pep	β/IS	pmol β	β/ protein	α/IS	pmol α	α/ protein	(α+β)/ protein
1	9,609	21,674	24,191	71.571	73.075	0.443	0.830	0.277	1.116	2.463	0.821	1.098
2	18,978	39,767	47,222	71.332	72.833	0.477	0.900	0.300	1.188	2.627	0.876	1.176
10	21,368	42,498	49,614	69.897	71.379	0.503	0.952	0.317	1.167	2.581	0.860	1.178
Average	15,255.2	33,306.5	39,176.5	72.033	73.543	0.456	0.855	0.285	1.174	2.596	0.865	1.150
SD	4,375.11	8,619.33	10,334.26	1.35	1.374	0.022	0.045	0.015	0.048	0.111	0.037	0.038

rules as well; however, we did not use the second internal standard peptide in these experiments. Future experiments are planned to test the effectiveness of this peptide as a potential internal standard.

Protein Quantification

The manually found standard curves and quantification results are taken from [9] and are used to verify the output of our system. The quantification results of patient samples are shown in Table 3.

Conclusions

Protein quantification becomes increasingly important for diagnoses and for elucidation of signal transduction pathways. Our system is designed to increase the efficiency of the quantification process as compared to manual quantification, which is labor intensive and prone to human error.

The key feature of our system is the integration of two subsystems into a single application for the discovery of quantification peptides, the design of the internal standard peptide, and the rapid and efficient quantification of the given data.

In short, we developed and validated a fully automatic protein quantification system. It is possible to use the system for similar problems.

Chia-Yu Yen is with the University of Colorado at Denver and Health Sciences Center.

Steve M. Helmke is with the University of Colorado at Denver and Health Sciences Center.

Krzysztof J. Cios is with the University of Colorado at Denver and Health Sciences Center, the University of Colorado at Boulder, and 4cData in Golden, Colorado.

M. Benjamin Perryman is with the University of South Dakota and Sioux Valley Hospitals and Health Systems.

Mark W. Duncan is with the University of Colorado at Denver and Health Sciences Center.

Address for Correspondence: Krzysztof J. Cios, University of Colorado at Denver and Health Sciences Center, Campus Box 109, Denver, CO 80217 USA. Phone: 303-556-6035. E-mail: Krys.Cios@cudenver.edu.

References

[1] G. Chambers, L. Lawrie, P. Cash, and G.I. Murray, "Proteomics: A new approach to the study of disease," *J. Pathol.*, vol. 192, no. 3, pp. 280–288, 2000.

[2] S.D. Patterson and R.H. Aebersold, "Proteomics: The first decade and beyond," *Nat. Genet.*, vol. 33, no. 35, pp. 311–323, 2003.

[3] B.D. Lowes, W. Minobe, W.T. Abraham, M.N. Rizeq, T.J. Bohlmeyer, R.A. Quaife, R.L. Roden, D.L. Dutcher, A.D. Robertson, N.F. Voelkel, D.B. Badesch, B.M. Groves, E.M. Gilbert, and M.R. Bristow, "Changes in gene expression in the intact human heart. Downregulation of alpha-myosin heavy chain in hypertrophied, failing ventricular myocardium," *J. Clin. Invest.*, vol. 100, no. 9, pp. 2315–2324, 1997.

[4] K. Nakao, W. Minobe, R. Roden, M.R. Bristow, and L.A. Leinwand, "Myosin heavy chain gene expression in human heart failure," *J. Clin. Invest.*, vol. 100, no. 9, pp. 2362–2370, 1997.

[5] B.D. Lowes, E.M. Gilbert, W.T. Abraham, W.A. Minobe, P. Larrabee, D. Ferguson, E.E. Wolfel, J. Lindenfeld, T. Tsvetkova, A.D. Robertson, R.A. Quaife, and M.R. Bristow, "Myocardial gene expression in dilated cardiomyopathy treated with beta-blocking agents," *N. Engl. J. Med.*, vol. 346, no. 18, pp. 1357–1365, 2002.

[6] S.P. Gygi, Y. Rochon, B.R. Franza, and R. Aebersold, "Correlation between protein and mRNA abundance in yeast," *Mol. Cell. Biol.,* vol. 19, no. 3, pp. 1720–1730, 1999.

[7] C.G. dos Remedios, D.A. Berry, L.K. Carter, J.V.F. Coumans, M.Y. Heinke, P.C. Kiessling, R.K. Seeto, T. Trahair, and M. Yao, "Different electrophoretic techniques produce conflicting data in the analysis of myocardial samples from dilated cardiomyopathy patients," *Electrophoresis*, vol. 17, no. 1, pp. 235–238, 1996.

[8] S. Miyata, W. Minobe, M.R. Bristow, and L.A. Leinwand, "Myosin heavy chain isoform expression in the failing and nonfailing human heart," *Circ. Res.,* vol. 86, no. 4, pp. 386–390, 2000.

[9] S.M. Helmke, C. Yen, K.J. Cios, K. Nunley, M.R. Bristow, M. Duncan, and M.B. Perryman, "Simultaneous quantification of human cardiac alpha- and beta-myosin heavy chain proteins by MALDI-TOF mass spectrometry," *Anal. Chem.,* vol. 76, no. 6, pp. 1683–1689, 2004.

[10] M. Bucknall, K.Y. Fung, and M.W. Duncan, "Practical quantitative biomedical applications of MALDI-TOF mass spectrometry," *J. Am. Soc. Mass. Spectrom,* vol. 13, no. 9, pp. 1015–1027, 2002.

[11] K.J. Cios and G.W. Moore, "Uniqueness of medical data mining," *Artif. Intell. Med.,* vol. 26, no. 1–2, pp. 1–24, 2002.

[12] K.J. Cios, W. Pedrycz, and R. Swiniarski, *Data Mining Methods for Knowledge Discovery*. Norwell, MA: Kluwer, 1998.

[13] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, and D.J. Lipman, "Basic local alignment search tool," *J. Mol. Biol.,* vol. 215, no. 3, pp. 403–410, 1990.

[14] S.B. Needleman and C.D. Wunsch, "A general method applicable to the search for similarities in the amino acid sequences of two protein," *J. Mol. Biol.*, vol. 48, no. 3, pp. 443–453, 1970.

[15] T.F. Smith and M.S. Waterman, "Identification of common molecular subsequences," *J. Mol. Biol.,* vol. 147, no. 1, pp. 195–197, 1981.

[16] M.O. Dayhoff, P.M. Schwartz, and B.C. Orcutt, "A model of evolutionary change in proteins," in *Atlas of Protein Sequence and Structure*, vol. 5, suppl. 3, M.O. Dayhoff, Ed. Washington, D.C.: National Biomedical Research Foundation, 1978, pp. 345–352, .

[17] D.C. Montgomery, E.A. Peck, and G. Vining, *Introduction to Linear Regression Analysis*, 3rd ed. New York: Wiley-Interscience, 2001.

[18] H. Motulsky and A. Christopoulos, *Fitting Models to Biological Data Using Linear and Nonlinear Regression,* 1999 [Online]. Available: http://www.curvefit.com/oldindex.htm