The changing face of newborn screening: diagnosis of inborn errors of metabolism by tandem mass spectrometry

Patricia M. Jonesa,*, Michael J. Bennetta,b

a Department of Pathology, University of Texas Southwestern Medical Center and Children's Medical Center, 1935 Motor Street, Dallas, TX 75235, USA
b Department of Pediatrics, University of Texas Southwestern Medical Center and Children's Medical Center, 1935 Motor Street, Dallas, TX 75235, USA

Received 8 April 2002; received in revised form 17 June 2002; accepted 17 June 2002

Keywords: Tandem mass spectrometry; Newborn screening; Inborn errors of metabolism

1. Introduction

Technological advances are revolutionizing the way in which we can diagnose and screen whole populations for inborn errors of metabolism. These advances are also necessitating a reevaluation of the spectrum of metabolic disorders included in newborn screening program. The initial concept for inclusion into a screening program was based upon a number of factors, including the use of one test on one sample for a single disease. Early assay capabilities limited the number of tests that a screening center could perform on the filter paper blood spot sample collected shortly after birth. As technology and assay sensitivity improved, more tests could be performed on a dried blood spot sample, expanding the laboratory’s ability to screen for more than one disorder in the newborn period. In the last decade, the evolution of tandem mass spectrometry has improved our diagnostic capabilities further. We can now readily detect multiple diseases on a single sample, using a single analytical procedure. This review will highlight historical aspects of newborn screening for inborn errors of metabolism, describe state-of-the-art screening practices and focus upon possible future trends for this important area of laboratory testing.

2. Inborn errors of metabolism

In the early 1900s, Garrod [1] coined the term “inborn errors of metabolism” to describe the hereditary deficiency or alteration in enzyme reactions and described the first “Inborn Error”, alkaptonuria. Before an understanding of metabolic pathways developed, diagnosis of inborn errors of metabolism was generally made by observation of gross abnormalities, such as the darkened urine in individuals with the clinical phenotype of alkaptonuria. Disorders were originally recognized in individuals who became severely ill, or who had dysmorphic features or mental aberrations, and these disorders were understood to be inherited by observation of recurrence in family members. With the identification of specific enzymes and metabolic pathways, our understanding of the physical processes causing these inherited metabolic disorders evolved. This knowledge was accompanied by the concomitant understanding of the possibility...
for errors at various stages in those pathways. It was recognized that deficient enzymes could be identified in individuals by analysis of body fluids and measuring the accumulation of substrate for that enzyme, or the accumulation of unusual products of pathways utilized in response to the enzymatic block.

At almost the same time as Garrod’s seminal observation, the basis of current concepts of genetic theory was being elucidated. Beadle and Tatum were instrumental in formalizing the concept of one gene–one enzyme. In this theory, an indivisible unit of heredity called a gene was thought to be responsible for the presence and activity of a single enzyme or protein [2,3]. The diagnostic search began for the altered or missing genes that produced the deficient enzymes, and an inborn error of metabolism was believed to be an “all-or-none” affair. Early on, only severe forms of inborn errors were diagnosed and it is only recently that we have come to understand that there is a continuum of mild to severe forms for almost every genetic defect known to date. Frequently, the first literature cases for inborn errors were those presenting with the severe phenotype, as these patients were easier to recognize in the population. Because the initial descriptions tend to define the phenotype for future patient investigations, milder variants often went undiagnosed.

The current state of knowledge allows the diagnosis of many inborn errors of metabolism by detection of the aberrant gene. However, as multiple genetic defects may result in the same biochemical phenotype, most diagnoses of genetic defects in a clinical setting are still accomplished by finding an abnormal accumulation of metabolites of the affected pathways, or unusual metabolites from alternate pathways. The advances made in the diagnosis of inborn errors of metabolism in recent years have not only been in the field of discovering new genes, but in technological advances in our ability to detect intermediates of metabolism in smaller samples, with greater efficiency and with fewer tests.

### 3. Newborn screening

In the early 1960s, Guthrie developed the first method applicable to whole population screening for phenylketonuria (PKU). In this assay, he measured phenylalanine in blood spots on filter paper using a bacterial inhibition assay [4,5]. In 1962, the state of Massachusetts instituted the first newborn screening program for PKU. At that point, it was already recognized that early detection of PKU and early introduction of a diet low in phenylalanine resulted in a significant decrease in morbidity, as well as significant savings of medical costs.

Ideally, detection of genetic disorders should occur very early in life, in a presymptomatic, well individual, before the metabolic error could disrupt mental or physical health and growth. The criteria for inclusion of a test/disorder into a newborn screening program continue to evolve. Currently the disorders included in most newborn screening programs essentially meet the following criteria:

1. have a significant incidence in the population screened,
2. are clinically well defined with the untreated natural history characterized,
3. have a well-defined biochemical phenotype,
4. cause significant morbidity and/or mortality,
5. are treatable, where treatment improves outcome,
6. testing is safe, simple and sufficiently sensitive.
7. specific confirmatory testing is available
8. testing, treatment and treatment outcome are cost effective with respect to non-treatment.

PKU serves as an excellent example of a metabolic disorder that fulfills these criteria [6]. The incidence at about 1 in 14,000 is sufficiently high to make PKU a significant public health issue. The biochemical phenotype is well defined, as is the untreated natural history in which patients follow a severe neurodegenerative course. When a patient with PKU is treated, the outcome is essentially that of a normal individual. Testing procedures are clearly cost effective. Currently, the cost of maintaining an individual with untreated PKU at an appropriate care facility is on the order of US$80,000/year. Fees charged for newborn screening range from US$15 to US$40 per sample screened, depending on the State and the level of screening offered. The actual cost to run just the PKU screen is probably in the neighborhood of US$4 per sample. Add to that the cost of follow up of individuals with positive screens and the cost of a low phenylalanine diet (~ US$6500/year for formula and
low protein diet), and the resulting figures are still impressively less than US$80,000/year.

By way of contrast with PKU, examples of disorders included in newborn screening programs that do not readily fit this list of criteria can be found easily. Table 1 gives the list of disorders that are screened for in the USA [7,8]. Of these disorders, Maple Syrup Urine Disease (MSUD) has a frequency in the general population of around 1 in 185,000, and galactosemia has an incidence of approximately 1 in 44,000 to 1 in 80,000. In the case of galactosemia, early treatment has not necessarily improved long-term outcome. Both of these diseases often present as a clinical crisis in the newborn period before the results of newborn screening tests are returned. Such disorders still require clinical diagnosis during the acute crisis in many instances. At present, newborn screening for galactosemia is performed in 48 of 50 states. Cystic fibrosis, with an incidence of approximately 1 in 2500, is presently screened for in only four states, with pilot programs in three more. Testing for the most common DNA mutation for this disorder only includes approximately 75% of affected individuals, causing this test to be considered insufficiently sensitive for inclusion in a general screening program. Medium-chain acyl-CoA dehydrogenase (MCAD) deficiency is just beginning to be included on newborn screens in some testing centers (seven states and one trial), and yet it has a relatively high incidence (approximately 1 in 14,000), has a significant mortality (25%) upon first presentation and is eminently treatable. Currently, we are undergoing a paradigm shift in choosing suitable criteria for disease inclusion in newborn screening programs, as new technologies make it possible to easily test for more disorders without additional cost. The leading technology driving these changes is the introduction of the tandem mass spectrometric methods for metabolite analysis.

### 4. Tandem mass spectrometry

Mass spectrometry (MS) is an ideal tool for the identification of metabolites. The very specific mass fragmentation patterns provide “chemical fingerprints”, and the only true positive identification mechanism for chemical compounds and metabolic intermediates. Chromatographic techniques are ideal for separation of compounds out of a complex mixture in a biological matrix, i.e. serum, but in the absence of MS, these techniques have always suffered from the problem of basing compound identification on retention time. Positive identification can never be made in a system where an unknown can co-elute at a known retention time. Thus, when mass spectrometry was introduced, coupled to gas chromatography with its high-efficiency separations, an ideal balance was achieved. Positive identification of compounds could now be made, based on very specific mass spectra, or ion fragment patterns. The drawbacks to this technique often involved the inability to make a compound volatile enough to separate it by gas chromatography. In addition, the long separation times, up to 65 min per sample for a urine organic acid analysis, make this analytical approach unsuitable for whole population screening. In recent years, mass spectrometers have been linked to liquid chromatographic systems. Liquid chromatography separates the compounds and mass spectrometry serves as the detector.

Table 1

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Approximate incidence</th>
<th>Number of states screening</th>
</tr>
</thead>
<tbody>
<tr>
<td>PKU</td>
<td>1:14,000</td>
<td>50</td>
</tr>
<tr>
<td>Congenital hypothyroidism</td>
<td>1:4000</td>
<td>50</td>
</tr>
<tr>
<td>Galactosemia</td>
<td>1:44,000–80,000</td>
<td>48</td>
</tr>
<tr>
<td>Sickle cell disease</td>
<td>1:40,000</td>
<td>47</td>
</tr>
<tr>
<td>Congenital adrenal hyperplasia</td>
<td>1:10,000–18,000</td>
<td>26</td>
</tr>
<tr>
<td>Biotinidase deficiency</td>
<td>1:60,000</td>
<td>25</td>
</tr>
<tr>
<td>Maple syrup urine disease</td>
<td>1:185,000</td>
<td>25</td>
</tr>
<tr>
<td>Homocystinuria</td>
<td>1:200,000</td>
<td>16</td>
</tr>
<tr>
<td>MCAD</td>
<td>1:14,000</td>
<td>8</td>
</tr>
<tr>
<td>Cystic fibrosis</td>
<td>1:2500</td>
<td>7</td>
</tr>
<tr>
<td>Tyrosinemia</td>
<td>1:100,000–120,000</td>
<td>2</td>
</tr>
<tr>
<td>Glucose-6-phosphate dehyd.</td>
<td>1:34–200</td>
<td>1</td>
</tr>
</tbody>
</table>

* Depending on population.

* 1:176 in Amish/Mennonite populations in Pennsylvania.

* 1:1846 in parts of Quebec, Canada.

* Incidence in males in the US. Incidence is 1:4–10 in parts of Africa and Asia.
the instrument to scan for either a precursor ion mass, a product ion mass after fragmentation of the compound, or a combination of the two. An excellent review of mass spectrometry methods has previously been written [9].

Tandem MS instruments have the capability of being used with a direct injection, when no separation is necessary, or being linked to a liquid chromatography system if some separation is required. In either case, run times for a tandem MS instrument are often

<table>
<thead>
<tr>
<th>Class</th>
<th>Disorder</th>
<th>Estimated incidence</th>
<th>Treatment improves outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino acids</td>
<td>Homocystinuria</td>
<td>1:200,000</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>Hypermethioninemia (1:20,000 in Denmark)</td>
<td>1:335,000</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>MSUD</td>
<td>1:185,000</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>PKU</td>
<td>1:14,000</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>Tyrosinemia, type 1 and 2</td>
<td>1:100,000–120,000</td>
<td>yes</td>
</tr>
<tr>
<td>Organic acids</td>
<td>GA1</td>
<td>1:50,000</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>GA2</td>
<td>not established&lt;sup&gt;a&lt;/sup&gt;</td>
<td>+/- &lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>HMG CoA lyase deficiency</td>
<td>&gt;100 cases&lt;sup&gt;c&lt;/sup&gt;</td>
<td>+/- &lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Isovaleric acidemia</td>
<td>&gt;70 cases</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>MCC deficiency</td>
<td>&gt;50 cases</td>
<td>+/-</td>
</tr>
<tr>
<td></td>
<td>MMA&lt;sup&gt;d&lt;/sup&gt;</td>
<td>1:48,000–61,000</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>PA</td>
<td>1:87,000–137,000&lt;sup&gt;f&lt;/sup&gt;</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>3-ketothiolase deficiency</td>
<td>&gt;50 cases</td>
<td>yes</td>
</tr>
<tr>
<td>Urea cycle</td>
<td>Argininaemia</td>
<td>1:363,000</td>
<td>+/- &lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>ASA</td>
<td>1:70,000</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>Citrullinemia</td>
<td>1:57,000</td>
<td>+/- &lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Fatty acids</td>
<td>CPT 1 deficiency</td>
<td>~ 20 cases&lt;sup&gt;f&lt;/sup&gt;</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>CPT 2 deficiency</td>
<td>&gt;200 cases&lt;sup&gt;f&lt;/sup&gt;</td>
<td>+/- &lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>C/AT defect</td>
<td>~ 10 cases&lt;sup&gt;f&lt;/sup&gt;</td>
<td>+/- &lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>LCHAD/TFP deficiency</td>
<td>&gt;200 cases&lt;sup&gt;g&lt;/sup&gt;</td>
<td>+/- &lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>MCAD deficiency</td>
<td>1:14,000</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>SCAD deficiency</td>
<td>not established&lt;sup&gt;d&lt;/sup&gt;</td>
<td>+/- &lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>VLCAD deficiency</td>
<td>&gt;200 cases&lt;sup&gt;f&lt;/sup&gt;</td>
<td>+/- &lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Other</td>
<td>Peroxisomal disorders</td>
<td>1:50,000</td>
<td>+/- &lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Lysosomal disorders</td>
<td>1:8000–200,000&lt;sup&gt;b&lt;/sup&gt;</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>Purine/pyrimidine disorders</td>
<td>not established</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>Steroid disorders&lt;sup&gt;i&lt;/sup&gt;</td>
<td>1:10,000–18,000</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>Bile acid disorders</td>
<td>not established</td>
<td>?</td>
</tr>
</tbody>
</table>

ASA = Argininosuccinic aciduria, CPT = carnitine palmitoyltransferase, C/AT = carnitine/acylcarnitine translocase, GA1, GA2 = glutaric acidemia, type 1 and type 2, HMG = 3-hydroxy-3-methylglutaryl, LCHAD = long-chain i-3-hydroxyacyl CoA dehydrogenase, MCC = 3-methylcrotonyl CoA carboxylase, MMA = methylmalonic acidemia, MSUD = maple syrup urine disease, PA = propionic acidemia, SCAD = short-chain acyl-CoA dehydrogenase, TFP = trifunctional protein, VLCAD = very-long-chain acyl-CoA dehydrogenase.

<sup>a</sup> Incidence unknown, but not rare. May be as common as GA1.
<sup>b</sup> Outcome depends on age of onset and severity.
<sup>c</sup> Greater than expected numbers being identified in tandem MS screening programs.
<sup>d</sup> Not all cases of MMA are detected by tandem MS screening, including false negatives in detecting the cobalamin defects.
<sup>e</sup> Refs. [31–33].
<sup>f</sup> Authors’ personal experience.
<sup>g</sup> Less than expected numbers being identified in tandem MS screening programs.
<sup>h</sup> Estimated incidence of all lysosomal disorders is ~ 1:8000; specifically for Pompe’s disease which is detectable by tandem MS, incidence is ~ 1:200,000 [52,53].
<sup>i</sup> Specifically 21-hydroxylase deficiency.
significantly shorter than HPLC or GC/MS run times. For comparison purposes, a normal GC/MS run for organic acids takes about 1 h. An ion-selective chromatographic method for amino acids can take anywhere from 30 min for a single phenylalanine measurement to 135 min for a screen for multiple amino acids. A phenylalanine assay by tandem MS takes less than 2 min. The rapidity of this assay is not its only positive asset in terms of newborn screening. The addition of tyrosine and use of a phenylalanine/tyrosine ratio has been shown to result in significantly fewer false positives, enhancing the utility of the assay [10]. Tandem MS is a very specific, very sensitive technology that has been utilized to detect a wide variety of analytes and diseases [11–21]. Table 2 shows some of the disorders that can currently be screened for by tandem mass spectrometry. In Table 2, the columns showing estimated incidence and treatment outcome are derived from the individual chapters for each disorder in The Metabolic and Molecular Bases of Inherited Disease, 8th ed., Charles R. Scriver, et al., editors [22], unless otherwise specified in the table. Using tandem MS, a single blood test is now capable of detecting more than 20 inborn errors of metabolism [23,24]. The technology and methodology are currently being scrutinized and recommendations made for utilization and standardization [25–29].

The disadvantages of MS/MS are those found with any new technology. The initial cost of a system is high. A level of expertise is needed for preparing samples and operating the system. And a level of expertise is required to interpret the data produced by the MS/MS. Many of the metabolic intermediates detected by this methodology are elevated in multiple different disorders, and so proficiency in interpretation is necessary. As with any screening test, follow-up testing for diagnosis is mandatory.

5. Results of ongoing screening programs

Table 3 summarizes the findings from five newborn screening programs that utilize tandem MS technology [30–34]. The studies presented ranged from 7 months to 7 years in length and 1100 to >700,000 patient samples tested. The studies clearly demonstrate the multiple utilities of this methodology. Not only are inborn errors of metabolism being diagnosed in a timely manner which allows for rapid treatment, tandem MS is also assisting in determining population incidence of various metabolic disorders, and diagnosing disorders that were previously being missed. In the large studies, the incidence of MCAD detected is between 1:14,000 and 1:20,000, except in Australia where it is 1:68,000. Prior to screening by tandem MS, 3-methylcrotonyl CoA carboxylase (MCC) deficiency was believed to be very rare, with only a few cases reported. However, in three of the large studies, MCC deficiency was diagnosed at incidences of 1:29,000 to 1:257,000, depending on the population [31–33]. The approximate incidences of the major types of disorders detected (amino acid, organic acid, etc.) are similar among studies, with the exception of the small study performed in Turkey. The Turkish study illustrates the much higher incidence rates that can be found with a small sample size and a specific population. Of interest in the large studies is the fact that the average overall incidence of a detectable inborn error of metabolism occurring is approximately 1 in every 4600 live births (New South Wales—1:4419; New England—1:5840; Neo Gen—1:4399; Bavaria—1:3954). As can be seen in Table 2, the clinical outcomes for the majority of these disorders are improved by treatment if diagnosis is made early and treatment is started quickly enough.
6. Future directions

With the advent of new technologies and the elucidation of new inborn errors of metabolism, the criteria for inclusion of a specific test or disease in a newborn screen are constantly being reconsidered [35]. An increasing number of reports in the recent literature are employing tandem MS to screen for and to diagnosis many different types of inborn errors of metabolism, including congenital disorders of glycosylation, disorders of purine and pyrimidine metabolism, adrenal cortical insufficiency, and galactosemia [36–39]. Tandem MS is being utilized for diagnosing disorders of fatty acid oxidation and related pathways, such as carnitine palmitoyltransferase II deficiency, very long chain acyl-CoA dehydrogenase deficiency and mitochondrial trifunctional protein deficiency [40–42]. New disorders are also being discovered with this technology [43,44]. Milder variants of diseases whose diagnosis may have been missed using older technologies are being recognized by newborn screening with tandem MS. A previously unrecognized genotype for MCAD deficiency has been identified by the ongoing screening programs [45,46]. Tandem mass spectrometry analysis of bile acids may be a useful diagnostic aid for other disorders, including peroxisomal disorders [47] and cholestatic hepatobiliary disease [48]. Cases such as the above suggest that tandem mass spectrometry may cause rapid expansion of the newborn screening programs, and cause us to rethink our criteria for newborn screening.

Simple tests are becoming available for rapid diagnosis of inborn errors of metabolism which have no known treatment that improves outcome, or are extremely rare. Both of these characteristics would have excluded these disorders from newborn screening in the past. Diagnosing these disorders whenever possible, however, can have an important effect on such areas as genetic counseling and new treatment possibilities. The question arises as to whether a test for an extremely rare disease should be routinely done simply because the test is easy to perform, inexpensive and available. When treatment could significantly alter outcome, there is very little debate over whether a disorder should be included in a newborn screen, unless the test is not specific enough or extremely costly. At some point this debate leaves cost considerations and becomes ethical. With the current ability of tandem MS to pick up elevated concentrations of metabolites and unusual metabolites, new and presently untreatable disorders will be detected, and the question becomes whether to report them. Although the controversy will continue over whether diagnosis of certain disorders should be done simply because it can be done, in the long run lack of knowledge may turn out to be more harmful than knowledge.

7. Summary

Newborn screening for inborn errors of metabolism is a field that is undergoing a revolution in all of its aspects [49,50]. Technological advances are driving our ability to assay for more disorders with little increase in time, resources or cost. The computer age is also a driving force in this field, as the general public becomes aware of new testing capabilities through magazines and Internet information flow and places its own demands on test menu and accessibility [51]. These rapid changes and increases in capabilities are in turn driving an evolving focus as to which disorders should be screened. Along with these alterations comes the responsibility to manage the changes appropriately. There is a growing need for standardization of all phases of the newborn screening program, including test menu and methodology, cutoff values, reporting of results and follow-up of positive results.

With the advances occurring in tandem mass spectrometric methodology, we predict that essentially all newborn screening for inborn errors will be done by this method within the next 10 years. Clinical laboratories performing this testing will need to be capable of and responsible for all aspects of the programs.

References


[34] Biberoğlu G, Hasanoglu A, Tumer L, Ezgu FS, Gene B. Screening inborn errors of metabolism from dried blood spots


