

CEPH Family 1362 STR Database: An Online Resource for Characterization of PCR Products Using Electrospray Ionization Mass Spectrometry

Allison P. Null and David C. Muddiman^{*,†}

Department of Chemistry, Virginia Commonwealth University, Richmond, Virginia, USA

An online database has been established in order to validate electrospray ionization mass spectrometry (ESI-MS) for genotyping and to publicize the procedures developed in our laboratory for the characterization of PCR products by ESI-MS. Genotypes derived from short tandem repeat (STR) loci that were obtained using ESI Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) have been posted for fifteen members of the CEPH family 1362 pedigree. The website provides specific information such as PCR parameters, PCR product cleanup approaches, and ESI solution compositions to enable other laboratories to reproduce our data. Links are provided to related websites in an effort to integrate information regarding the CEPH family, STR genotyping, and mass spectrometry. The database, currently available at <http://www.people.vcu.edu/~dcmuddim/genotype/> will be routinely updated with genotypes from additional STR loci including PCR parameters as well as PCR cleanup strategies as further developments are completed. (J Am Soc Mass Spectrom 2002, 13, 89–90)
© 2002 American Society for Mass Spectrometry

The central theme of the post genome era will be elucidating the association of genetic variation with specific disease phenotypes. This approach requires the screening of numerous genetic markers to narrow the potential locations of a disease gene. The most useful of these genetic markers are short tandem repeats (STRs), which are highly polymorphic repeat motifs found every 10–20 kb in the human genome [1–3]. In addition to their utility as genetic markers, some STRs are directly linked to a class of hereditary disorders (i.e., Huntington's disease, spinocerebellar ataxia [4]), and are utilized for human identification in forensic science [5]. ESI-MS [6], specifically FTICR-MS [7, 8], is emerging as a significant player in the accurate, sensitive, high-throughput measurement of STR length polymorphisms [9]. FTICR presents advantages over other types of mass detectors for characterization of biomolecules because of its unparalleled mass resolution on the order of 10^5 or better [10–15] and superior limit of detection allowing spectra of PCR products to be generated from 5 nM solutions [16].

Although ESI-MS has been shown to be a versatile

and robust approach for characterization of PCR products [9, 16–30], it is still in its infancy. To validate ESI-MS as a tool for characterizing PCR products and to make available the approach that has consistently proven successful in our laboratory, we have established the ESI-FTICR-MS database for CEPH family 1362. This database is currently online and available at <http://www.people.vcu.edu/~dcmuddim/genotype/>. Should the URL change during the course of this project, a virtual redirect will be implemented. The database is intended to provide model systems for STR genotyping and relevant details that should enable others to develop mass spectrometry for routine analysis of nucleic acids. In addition, it is part of our dissemination of information to the broader mass spectrometry community as outlined in our National Human Genome Research Institute grant.

Scope

All CEPH family genotypes to date were obtained using ESI-FTICR-MS and the genotype data were confirmed using other resources (e.g., websites, publications) when available. Genotypes can easily be searched from the homepage by chromosome, STR locus or family member. If searching by chromosome, a map of the selected chromosome is displayed with all available loci listed in their relative positions. When a locus is se-

Published online November 27, 2001

Address reprint requests to Dr. D. C. Muddiman, Department of Chemistry, Virginia Commonwealth University, 1001 West Main Street, P.O. Box 842006, Richmond, VA 23284, USA. E-mail: dcmuddim@mail1.vcu.edu

^{*}Affiliate appointment in Biochemistry and Molecular Biophysics.

[†]The Massey Cancer Center.

lected from a particular chromosome or from an alphabetical list of all loci, the genotype for each member of the CEPH family 1362 is shown in pedigree format. Alternatively, the user can choose a specific family member; a page for each family member displays the genotype for each available STR locus.

In addition to the genotypes, detailed information is provided in this database regarding the methods that are utilized for successful analysis of PCR products by ESI-MS. The PCR parameters for each STR locus are provided in PDF format, which can be accessed from the homepage or from the subsequent pages for each locus. Primers, reagents, reaction composition, and cycling conditions are detailed for each locus. The post PCR cleanup procedure that is used by our laboratory consists of an ethanol precipitation followed by microdialysis, both of which are detailed at this site. We are, of course, continually in search of a more rapid and efficient approach to sample cleanup. As alternative approaches are elucidated, they will be added to the database. ESI solution compositions used to generate both single- and double-stranded PCR products by ESI-FTICR are given. Finally, for each STR locus, lists of average masses and molecular formulas are provided for the single- and double stranded PCR products for each known allele.

The ESI-FTICR mass spectrometry genotype database for CEPH family 1362 will be routinely maintained to add new genotypes and new STR loci as the data are acquired. Due to recent successes with ESI time-of-flight (TOF) mass spectrometry, which is a potentially higher throughput technology than FTICR, we anticipate addressing simpler STRs with ESI-TOF and adding these data to the database.

Conclusions

It is our intention to provide mass spectrometrists with information which will facilitate the development and use of mass spectrometry for characterizing PCR products. STRs have been chosen as a model system not only because of their importance in genetics, but because they vary in size and complexity. For information not included, links to other websites are provided, thus establishing a network of information for STRs, CEPH families, and mass spectrometry. We encourage users to provide feedback and make suggestions as to what could improve the database. It is recommended that this paper be cited in publications rather than the URL as the latter may change periodically.

Acknowledgments

The authors gratefully acknowledge the financial support of the National Institutes of Health (R01HG02159) and the Mary E. Kapp Foundation of the Department of Chemistry, Virginia Commonwealth University.

References

- Litt, M.; Luty, J. A. *Am. J. Hum. Genet.* **1989**, *44*, 397–401.
- Tautz, D. *Nucleic Acids Res.* **1989**, *17*, 6463–6471.
- Edwards, A.; Civitello, A.; Hammond, H. A.; Caskey, T. C. *Am. J. Hum. Genet.* **1991**, *49*, 746–756.
- Wells, R. D.; Warren, S. T. *Genetic Instabilities and Hereditary Neurological Diseases*. Academic Press: San Diego, 1998, pp 3–12.
- Alford, R. L.; Caskey, C. T. *Curr. Opin. Biotechnol.* **1994**, *5*, 29–33.
- Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64–71.
- Comisarow, M. B.; Marshall, A. G. *Chem. Phys. Lett.* **1974**, *25*, 282–283.
- Henry, K. D.; Williams, E. R.; Wang, B. H.; McLafferty, F. W.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 9075–9078.
- Null, A. P.; Muddiman, D. C. *J. Mass Spectrom.* **2001**, *36*, 589–606.
- McLafferty, F. W. *Acc. Chem. Res.* **1994**, *27*, 379–386.
- Marshall, A. G.; Senko, M. W.; Li, W. Q.; Li, M.; Dillon, S.; Guan, S. H.; Logan, T. M. *J. Am. Chem. Soc.* **1997**, *119*, 433–434.
- Wu, Q. Y.; Vanorden, S.; Cheng, X. H.; Bakhtiar, R.; Smith, R. D. *Anal. Chem.* **1995**, *67*, 2498–2509.
- Speir, J. P.; Senko, M. W.; Little, D. P.; Loo, J. A.; McLafferty, F. W. *J. Mass Spectrom.* **1995**, *30*, 39–42.
- He, F.; Hendrickson, C. L.; Marshall, A. G. *Anal. Chem.* **2001**, *73*, 647–650.
- Bowers, M. T.; Marshall, A. G.; McLafferty, F. W. *J. Phys. Chem.* **1996**, *100*, 12897–12910.
- Hannis, J. C.; Muddiman, D. C.; Fresenius, J. *Anal. Chem.* **2001**, *369*, 246–251.
- Muddiman, D. C.; Wunschel, D. S.; Liu, C. L.; Pasatolic, L.; Fox, K. F.; Fox, A.; Anderson, G. A.; Smith, R. D. *Anal. Chem.* **1996**, *68*, 3705–3712.
- Naito, Y.; Ishikawa, K.; Koga, Y.; Tsuneyoshi, T.; Terunuma, H. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 1484–1486.
- Wunschel, D. S.; Fox, K. F.; Fox, A.; Bruce, J. E.; Muddiman, D. C.; Smith, R. D. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 29–35.
- Wunschel, D. S.; Muddiman, D. C.; Fox, K.; Fox, A.; Smith, R. D. *Anal. Chem.* **1997**, *70*, 1203–1207.
- Hannis, J. C.; Muddiman, D. C. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 954.
- Muddiman, D. C.; Null, A. P.; Hannis, J. C. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 1201–1204.
- Null, A. P.; Hannis, J. C.; Muddiman, D. C. *The Analyst* **2000**, *125*, 619–625 Special Issue on Biological Mass Spectrometry.
- Hannis, J. C.; Muddiman, D. C.; Null, A. P. *Advances in Nucleic Acid and Protein Analyses, Manipulation, and Sequencing. Proc. SPIE*, San Jose, CA. **2000**, 3926, 36–37.
- Krahmer, M. T.; Johnson, Y. A.; Walters, J. J.; Fox, K. F.; Fox, A. F.; Nagpal, M. *Anal. Chem.* **1999**, *71*, 2893–2900.
- Hannis, J. C.; Muddiman, D. C. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 876–883.
- Hannis, J. C.; Muddiman, D. C. *Rapid. Commun. Mass Spectrom.* **2001**, *15*, 348–350.
- Hahner, S.; Schneider, A.; Ingendoh, A.; Mosner, J. *Nucleic Acids Res.* **2000**, *28*, e82.
- Null, A. P.; Hannis, J. C.; Muddiman, D. C. *Anal. Chem.* **2001**, *73*, 4514–4521.
- Krahmer, M. T.; Walters, J. J.; Fox, K. F.; Fox, A.; Creek, K. E.; Pirisi, L. *Anal. Chem.* **2000**, *72*, 4033–4040.