Online-Only Abstract

Updated model of group A *Streptococcus* M proteins based on a comprehensive worldwide study

D. J. McMillan[1], P.-A. Drèze[2], T. Vu[1], D. E. Bessen[3], J. Guglielmini[4,5], A. C. Steer[6,7], J. R. Carapetis[8], L. Van Melderen[2], K. S. Sriprakash[1] and P. R. Smeesters[1,2,7] for The M Protein Study Group

1) *Bacterial Pathogenesis Laboratory, Queensland Institute of Medical Research, Brisbane, Qld, Australia,* 2) *Laboratoire de Génétique et Physiologie Bactérienne, Institut de Biologie et de Médecine Moléculaires, Faculté des Sciences, Université Libre de Bruxelles, Belgium,* 3) *Department of Microbiology and Immunology, New York Medical College, Valhalla, NY, USA,* 4) *Département Génomes et Génétique, Microbial Evolutionary Genomics, Institut Pasteur,* 5) *CNRS, UMR3525, Paris, France,* 6) *Centre for International Child Health, Department of Paediatrics, University of Melbourne, Royal Children's Hospital, Melbourne, Victoria,* 7) *Murdoch Children Research Institute, Melbourne, Victoria and* 8) *Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia, Perth, WA, Australia*

Original Submission: 19 September 2012; **Revised Submission:** 2 December 2012; **Accepted:** 14 December 2012
Editor: J.-L. Mainardi
Article published online: 20 December 2012
Clin Microbiol Infect 2013; **19:** E222–E229
10.1111/1469-0691.12134

Abstract

Group A *Streptococcus* (GAS) M protein is an important virulence factor and potential vaccine antigen, and constitutes the basis for strain typing (*emm*-typing). Although >200 *emm*-types are characterized, structural data were obtained from only a limited number of *emm*-types. We aim to evaluate the sequence diversity of near-full-length M proteins from worldwide sources and analyse their structure, sequence conservation and classification. GAS isolates recovered from throughout the world during the last two decades underwent *emm*-typing and complete *emm* gene sequencing. Predicted amino acid sequence analyses, secondary structure predictions and vaccine epitope mapping were performed using MUSCLE and Geneious software. A total of 1086 isolates from 31 countries were analysed, representing 175 *emm*-types. *emm*-type is predictive of the whole protein structure, independent of geographical origin or clinical association. Findings of an *emm*-type paired with multiple, highly divergent central regions were not observed. M protein sequence length, the presence or absence of sequence repeats and predicted secondary structure were assessed in the context of the latest vaccine developments. Based on these global data, the M6 protein model is updated to a three representative M protein (M5, M80 and M77) model, to aid in epidemiological analysis, vaccine development and M protein-related pathogenesis studies.