Abstract

Paul Piccinelli **Computational identification of non-coding RNAs**
Department of Medical Biochemistry and Cell Biology, Institute of Biomedicine, Sahlgrenska Academy at Göteborg University, Box 440, SE-405 30 Göteborg.

Problem. A large amount of genomic information is now becoming available. Suitable bioinformatic tools to organize and analyze this vast amount of information are therefore important. In the case of protein genes, the majority of these may be correctly identified using standard search methods that are based on sequence alignment. However, a different problem is presented when analysing non-coding RNA genes, since for their identification it is essential to take into consideration secondary structure features. Secondary structure is not only important for non-coding RNA genes, but it is also important in the regulation of gene expression. This work is concerned with the development of methods for ncRNA prediction and the application of these methods to identify specific ncRNA families.

Methods. Bioinformatic methods are used to identify ncRNA genes and ncRNA regulatory motifs. These methods include *de novo* methods, statistical profiles for primary sequence and secondary structure, sequence homology methods and minimum free energy methods. For protein gene identification we have used primary sequence alignments and profile searches and for protein classification we have used phylogenetic methods.

Results. RNase P and RNase MRP are two related ribonucleoprotein particles involved in RNA processing. We have used an approach based on conserved sequence elements to computationally analyze various eukaryotic genomic sequences for P and MRP RNA genes. We have found over 100 novel sequences, all able to fold into the consensus secondary structure of P and MRP RNAs. These genes reveal further evidence of the evolutionary relationship between these RNAs.

We also performed a computational analysis of the P/MRP protein subunits in eukaryotic organisms. A number of novel homologues were identified and we found novel orthologous relationships between fungal and metazoan proteins. Our results further emphasize a structural and functional similarity between the yeast and human P/MRP complexes.

The iron responsive element (IRE) is an RNA hairpin structure located in certain genes that are post-transcriptionally regulated in response to iron. We have found more than 90 novel sequences with the characteristics of known IREs. We have found evidence that the ferritin IRE represents the ancestral version of this type of translational control.

Finally, ncRNA genes in yeast have been predicted using two the *de novo* methods, RNAz and QRNA. A number of predicted candidates have been selected for experimental testing and more candidates will be tested.

Conclusions. We have used different bioinformatic methods to identify ncRNAs in a variety of organisms and report on several ncRNA sequences not previously reported. These novel RNA sequences make it possible to better predict the structure of these RNAs as well as to better understand their evolution and function. To further understand the structure and evolution of the RNases P and MRP we also analyzed the protein composition of these enzymes. Together, these new predictions aid to better understand the structure, function and evolution of RNase P and MRP.

Keywords: RNase P, RNase MRP, IRE, non-coding RNA, secondary structure
ISBN 978-91-628-7056-0 (91-628-7056-0) Göteborg 2006

Computational identification of non-coding RNAs

AKADEMISK AVHANDLING

som för avläggande av medicine doktorsexamen vid Sahlgrenska akademin vid Göteborgs universitet kommer att offentligen försvaras i föreläsningssal Arvid Carlsson, Medicinargatan 3A

Fredagen den 2 februari 2007, kl 09.00

av

Paul Piccinelli

Fakultetsopponent: Professor Jan Gorodkin, The Royal Veterinary and Agricultural University, Frederiksberg, Denmark

Handledare: Professor Tore Samuelsson, Institutionen för biomedicin

Avhandlingen baseras på följande arbeten:

I. Identification and analysis of ribonuclease P and MRP RNA in a broad range of eukaryotes
Paul Piccinelli, Magnus Alm Rosenblad, Tore Samuelsson.
Nucleic Acids Res. 2005 Aug 8;33(14):4485-95.

II. Inventory and analysis of the protein subunits of the ribonucleases P and MRP provides further evidence of homology between the yeast and human enzymes
Magnus Alm Rosenblad, Marcela Dávila López , Paul Piccinelli and Tore Samuelsson.
Nucleic Acids Res. 2006 Sep 34(18):5145-5156.

III. Evolution of the iron responsive element
Paul Piccinelli and Tore Samuelsson.
Submitted for publication.

IV. Hunting for non-coding RNA genes in yeast
Paul Piccinelli, Jonathan Esguerra, Anders Blomberg, Tore Samuelsson.
In manuscript.