Article

Combinatorial peptide library binding of mammalian spermatozoa identifies a ligand (HIPRT) in the axin protein: putative identification of a sperm surface axin binding protein and intriguing developmental implications

Dr Pieczenik is a notable graduate of Andover Academy, and studied at Harvard University. He worked with Fred Sanger, Francis Crick, Sydney Brenner, Bruce Merrifield, Norton Zinder and other early founders of molecular biology at the MRC Laboratory of Molecular Biology in Cambridge, UK and the Rockefeller Institute in New York, USA. He is considered the father of combinatorial chemistry and made the first combinatorial libraries at the MRC Laboratory of Molecular Biology. For the past several years, Dr Pieczenik has been involved, with his co-authors and others, in studies related to whole-cell screening of gametes and early embryos. This work has led to several intriguing findings some of which are presented in the current manuscript.

Dr George Pieczenik

Henry E Malter[1], Jacques Cohen[1,3], George Pieczenik[1,2]
[1]Tyho-Galileo Research Laboratories, 101 Old Short Hills Road, Suite 501, West Orange, NJ 07052; [2]Rutgers University, New Brunswick, NJ 08903, USA
[3]Correspondence: e-mail: jc@embryos.net

Abstract

The identification of components in cell–cell interactions is an important research goal in reproductive and developmental biology. Such interactions are critical to gamete development, fertilization, implantation and basic development. Several proteins involved with sperm–oocyte interaction and other developmentally important phenomena have been identified. However, these are obviously only a subset of the molecular components involved in such complex cell–cell interactions. One method that has been used to identify binding partners for particular molecular targets is the use of combinatorial libraries accessible on phage surfaces. For the most part, this technique has mainly been applied to screen specific target moieties. However, in some cases whole-cell screening has been attempted. This study describes the first report of screening intact, living mammalian gametes using a proprietary whole-cell combinatorial library binding and analysis protocol. Results from the first screening protocol of mouse spermatozoa strongly identified a putative sperm-binding ligand using proprietary bioinformatic analysis. This amino acid sequence (HIPRT) precisely corresponds with a previously characterized highly conserved protein–protein interaction site in the axin protein. This sequence is found within the binding site for a known sperm surface protein, glycogen synthase kinase-3. This result not only provides proof of the utility of this technique to identify cell surface ligands in mammalian gametes, but it also suggests a potential role for spermatozoa in facilitating developmental axis formation in mammalian embryos.

Keywords: *axin, combinatorial libraries, GSK-3, HIPRT, peptides, spermatozoa*

Introduction

Understanding cell–cell interactions is a key goal in many areas of cell and developmental biology. This is particularly so in the case of early development. Despite years of basic research into such interaction, much remains mysterious (Myles *et al*., 1994; Evans *et al*., 1995; McElhinny and Warner, 2000; Neganova *et al*., 2000; Ensslin and Shur, 2003). Elucidating the key components involved with cell-cell interactions in fertilization, cleavage, implantation and other phenomena will not only provide great insights into the basic developmental process, but will also be of great value in the development of interventional techniques in assisted reproduction and contraception. A variety of techniques have been used in analysing the components involved with these cell–cell phenomena. For the most part, these strategies have

attempted to define the cell surface proteome and determine the function and importance of specific isolated proteins via genetic or biochemical analysis. Ideally, the vital cell surface could be screened *in situ*, allowing for a more natural and complete understanding of the components involved in cell–cell interactions

One approach is provided through the use of combinatorial libraries. (Pasqualini and Ruoslahti, 1996; Szardenings *et al*., 1997; Pieczenik, 1999, 2003a,b; Merrifield, 2001; Cao *et al*., 2003). With the completion of the human genome sequence, the information for all the possible sequences involved in all interactions in the human is theoretically available. The question is how to sort out all the interactions. A peptide combinatorial library can contain all possible sequences in the range of four to seven amino acids. The great preponderance of non-covalent binding interactions at the molecular level are made of interactions between small number of amino acids in the range of four to seven or at extremes, multiples of such sizes. Another constraint is the number of unique peptide sequences in the genome. Even if the whole genome were coding, e.g. 3 billion nucleotides coding for a unique 1 billion amino acids, then any given sequence could be uniquely specified by six to seven amino acids. This can be seen in that 20 to the power 6–7 has a range of 64 million to 1.28 billion. In reality, there are only about 20,000 coding sequences in the genome and even if each coding is generously assigned 500 amino acid lengths, this makes about 10 million unique sequences. This database can therefore be uniquely specified or searched by five or six amino acids. That is, one would expect to see any given six amino acid sequence appear once in the human genome. The figure of five amino acids is considered the unique determinant for almost all known protein sequences. Therefore, both the physical chemistry of interactions and the global statistic of the genome codings suggest that five amino acids constitutes a unique peptide and protein identifier (Pieczenik, 2003b).

Therefore, the genome is effectively a closed informational circle for sequences in the range of four to seven amino acids. The universe of antibody–antigen interactions is also not infinite, but defined in size (Pieczenik, 2003b). This is a strategy that was used in sequencing the ΦX174 (Sanger and Coulson, 1975). The ΦX174 genome is actually a physical circle and therefore continuous random sequencing will effectively 'close the circle'. A combinatorial library containing all possible four-to-seven amino acid sequences essentially creates a 'virtual' closed circle of all possible interacting ligands for the expressed genome or protein phenotype. This aspect of the closed and limited size of both the genome codings and the combinatorial libraries guarantees that if a binding ligand is identified in the range of four to seven amino acids, then chances are that it is unique to the function of that specific binding.

Intact live mouse spermatozoa have been screened using a proprietary combinatorial peptide library selection protocol. Following the initial four rounds of binding, amplification and selection, a consistent set of peptide sequences emerged as a putative ligand interacting with sperm surface component. Upon further analysis, the convergent peptide sequence (HIPRT) was identified and shown to be present in several proteins, including a well characterized, completely conserved region of the axin protein mediating interaction w/glycogen synthase kinase-3 (GSK-3) (Zeng *et al*., 1997; Ikeda *et al*., 1998). GSK-3 has been recently identified as a mammalian sperm surface protein with a potential role in the regulation of motility (Vijayaraghavan *et al*., 1996, 2000; Smith *et al*., 1999; Somanath *et al*., 2004). This result demonstrates that whole cell combinatorial peptide library selection is a valid approach for the identification of sperm surface proteins. Furthermore, it provides intriguing molecular evidence supporting the theory that spermatozoa may play a role in the determination of polarity in mouse embryos.

Materials and methods

A mouse sperm suspension was isolated from cauda epididymides of fertile male CB6F1 mice. Briefly, the epididymides and vas deferens were dissected in a 35 mm dish containing 5 ml of CZB medium (Chatot *et al*., 1989) (pre-equilibrated at 37°C and 5% CO_2) and gently compressed to release spermatozoa. Following a 10-min incubation period, the epididimydes were removed and the sperm suspension incubated for a further 30-min period. A 500 µl aliquot of the sperm suspension was removed from the top layer of medium and placed in a 1.5 ml centrifuge tube. A 3 µl volume of phage combinatorial peptide library (see below) was added to this tube and the suspension incubated for 1 h at 37°C. Following incubation, the sperm suspension was washed as follows. The suspension was centrifuged in a tabletop microcentrifuge at maximum speed for 1 min and the majority of the supernatant removed. The remaining volume was re-suspended in 1 ml of medium via pipetting and separated into two 500 µl aliquots in two 15 ml centrifuge tubes. A further 14 ml of CZB wash medium was added to these tubes followed by 15 min incubation at room temperature. The tubes were centrifuged at 1000 *g* for 5 min and the supernatant removed down to an approximate 50 µl volume. A further 14 ml of CZB wash medium was added, followed by a 15 min incubation period at room temperature and identical centrifugation step. The supernatant was removed from each tube down to an approximate 50 µl volume and this remaining volume was combined with a further volume of CZB medium to a final volume of 500 µl for library isolation. At each step in the protocol, an aliquot was examined and shown to contain motile sperm cells. In addition, aliquots were taken from the original suspension and wash volumes for assessment of the efficiency of the screening and washing protocol. This process was repeated three times using the screened and amplified library created by prior screens. However, in the second to fourth screening protocols, a 50 µl volume of library was added to the 500 µl sperm suspension.

Combinatorial peptide library handling

The original combinatorial peptide library used in these experiments was derived from the 'PhD' Phage Display Peptide library (New England Biolabs, Beverly, MA, USA). General procedures and handling were done according to the protocols provided with these libraries. Bound phage were eluted at several different pH values in the range of 2–5 and several concentrations in the range of 0.1–1 mol/l Tris glycine at different times in the binding cycle. There was no panning, as described in the New England Biolab protocols.

Phage DNA was isolated by two different methods. One procedure used Qiagen protocols (Qiagen Inc., Valencia, CA, USA) and another procedure used AmpliTemp protocols