Kinesins in the Arabidopsis genome: A comparative analysis among eukaryotes

ABSTRACT

There are groups of Arabidopsis kinesins that are not present in yeast, Caenorhabditis elegans and Drosophila melanogaster that may, therefore, represent new subfamilies specific to plants. The domains present in different kinesins may provide clues about their functions in cellular processes. The comparative analysis presented here provides a framework for future functional studies with Arabidopsis kinesins.

INTRODUCTION

Background During the last five years, the genomes of four diverse eukaryotic organisms have been completely sequenced. These include Saccharomyces cerevisiae, a unicellular eukaryote, two multicellular organisms, Caenorhabditis elegans, and Drosophila melanogaster and Arabidopsis thaliana, the first plant genome to be completed. The sequencing of these major groups of organisms allows comparative analysis of genes, gene families, and genomes across phylogenetically divergent organisms. A study of predicted gene families in various organisms from single-cell to more complex animal and plant species is of value in deducing functions of proteins and developmental control pathways. Comparison of family members to proteins in its own genome and the genomes of other organisms can lead to further characterization of the proteins and provide clues to their function. With the information gained from these analyses, experimental procedures could be designed to determine the functions of the proteins. Budding yeast has 113 conventional protein kinase genes. Using multiple alignment and parsimony analysis of protein kinase catalytic domain sequences, the yeast protein kinases were categorized into subfamilies based on structural relatedness. This information led to identification of yeast specific kinases and gave hints about the function of unknown kinases. An analysis of proteins containing zinc finger domains in bacteria, yeast and C. elegans revealed that bacteria do contain some proteins that bind zinc but lack large families of zinc-binding proteins found in eukaryotes. Between eukaryotes, the presence and size of the different zinc-binding families vary. In a comparative analysis of the genomes of Drosophila, C. elegans and S. cerevisiae the number of proteins for each of the 200 most frequently occurring protein domains was identified for each organism. Among these domains were various calcium-binding domains such as the EF-hand family. In Drosophila there are 130 proteins with this domain while in C. elegans there are 79 and only 16 in yeast. A BLASTP and TBLASTN search of the Drosophila genome with members of the cytoskeleton protein families identified 262 genes with moderate to completely convincing homology to cytoskeletal genes. New members of the families were discovered and some proteins that are present in other genomes were missing in the Drosophila genome. Kinesins constitute a superfamily of microtubule (MT) motor proteins found in all eukaryotic organisms. Members of the kinesin superfamily have a highly conserved motor domain. The first kinesin was identified in squid giant axons as a protein involved in transport of vesicles. The conventional kinesin is a tetramer with two heavy chains and two light chains. The kinesin heavy chains (KHCs) contain the motor domain with ATPase activity, a central coiled-coil region and a tail that binds the light chains. Historically, proteins with homology to KHCs but falling in different subfamilies have been called kinesin-like proteins. However, KHCs are now recognized as a subfamily of

the kinesin superfamily and all members of the superfamily are referred to as kinesins. We have followed that pattern in this paper. KHCs have been identified in fungi and animals and a large number of other kinesins from other subfamilies have been identified in all eukaryotes. All kinesins have a domain with homology to the motor domain of KHC but little sequence similarity outside of this domain. Some kinesins have a coiled-coil region but others do not. The tail domain, which is believed to interact with specific cargoes, is nonconserved. Kinesin bind MTs and a variety of cargoes and perform force-generating tasks such as transport of vesicles and organelles, spindle formation and elongation, chromosome segregation and MT organization. The motor domain of KHC is in the N-terminal region but in other kinesins it can also be located in the C-terminus or internally. Motility assays have been performed with a number of the kinesins. The C-terminal domain kinesins have been shown to have minus-end motility while the others have plus-end motility. Nine subfamilies of kinesins have been identified by phylogenetic analysis using the conserved motor domain. Not all kinesins fall into one of the nine subfamilies and may represent additional subfamilies. The first two plant kinesins identified were found in tobacco pollen tube (pollen kinesin homologue, PKH) and tobacco phragmoplast (tobacco kinesin related protein, NtKRP125). Another kinesin isolated more recently from tobacco pollen tubes has also been characterized at the biochemical level. Kinesins that have been characterized at the molecular level in plants include, NtKRP125 in tobacco, four Arabidopsis kinesins identified by PCR-based cloning (KatA, KatB, KatC, KatD), KCBP (kinesin-like calmodulin binding protein) found in Arabidopsis, tobacco, potato, and maize, PAKRP in Arabidopsis and DcKRP120-1 and DcKRP120-2 in carrot. With the complete sequencing of the Arabidopsis genome, it has become possible to search the Arabidopsis database with the conserved motor domain of kinesins to identify the kinesins encoded in the Arabidopsis genome. We have identified 61 kinesin-like genes in Arabidopsis and their general location on the five Arabidopsis chromosomes. Surprisingly, the Arabidopsis genome contains the largest number of kinesins among all eukaryote genomes that have been sequenced. We have further analyzed the predicted protein sequences for the presence of domains that might lead to an understanding of their function. By definition all have a kinesin motor domain and several have a coiled-coil domain that might indicate dimerization. A phylogenetic analysis of the 61 Arabidopsis kinesin motor domain sequences with 113 other motor domain sequences has revealed that there are Arabidopsis kinesins that fall into seven of the nine recognized subfamilies of kinesins. Some do not fall into any family and there are some subfamilies unique to Arabidopsis and maybe to plants.

CONCLUSION

Conclusions In summary, Arabidopsis has a surprisingly large number of kinesins among the five completed eukaryotic genomes. Many Arabidopsis kinesins do not fall into any known subfamilies of kinesins and several Arabidopsis kinesins are not present in yeast, C.elegans and Drosophila and are likely to represent new subfamilies specific to plants. Further analyses of kinesins have resulted in identification of several interesting domains in Arabidopsis kinesins that provide clues in understanding their functions. Although the functions of most of the Arabidopsis kinesins remains to be determined, phylogenetic analysis of kinesins and identification of functional domains in these proteins provide clues to their function which can be tested empirically. Several knockout mutant libraries obtained by T-DNA insertions are available to screen for mutations in kinesins (). An alignment of the motor domains of the Arabidopsis kinesins

revealed a few very conserved stretches of amino acids in the motor domain that could be used for designing a universal degenerate primer set to screen for mutations in all kinesins. Once an insertion of T-DNA is detected in a kinesin, the sequence of the amplified product could be obtained to identify the kinesin involved. Due to the redundancy in function that has been seen with the non-plant kinesins, other strategies such as overexpression of the kinesins may also be needed. Protein-protein interaction studies using the yeast two-hybrid trap and expression analysis of all kinesins in different tissues and cell types with microarrays can also provide valuable information about the function of the kinesins.