Accessing and distributing EMBL data using CORBA (common object request broker architecture)

ABSTRACT

By integrating CORBA into existing systems and providing users with built-in clients for their applications, we can provide a more efficient way to access EMBL data while reducing the challenges of traditional flat-file formats.

INTRODUCTION

Many disciplines have developed graphical representations of complex systems to communicate, archive and analyze knowledge. Although electronic circuit diagrams and architectural plans are accessible to knowledgeable individuals, there is no universal'real' describing functional systems in biology. Knowledge in specific areas, such as metabolic pathways, gene networks, signaling networks and molecular interactions, is stored and archived in various forms in print or online, but these representations are not standardized to account for cross-disciplinary systems – for example, the interactions between genes and metabolism at the inter- and intra-molecular levels. Inadequate cartoon diagrams are frequently used by biologists to represent multidisciplinary biological problems at various levels, such as intramolecular and disease phenotype, due to their ambiguity and lack of clarity. We propose that a standardized visual biological description language would offer easier and more clear communication, and, with computational resources, provide underlying conditions for distributed searchable archives of functional knowledge, as well as CAD functionality for simulating and analyzing biological systems. Our team is currently developing a prototype biological description language, BioD, to test its conceptual foundation, explore its utility, and identify critical issues surrounding its implementation.

CONCLUSION

Annotations for a new genome are likely to be equally good as the database(s) to which they are compared. The COG database was established on the basis of the phylogenetic principle of protein classification, which involves clustering by (probable) orthology. Furthermore, significant efforts have been made to characterize and classify the COGs for functional purposes. This resulted in many benefits for annotating new genomes of organisms outside of well-characterized groups, including obtaining important functional predictions that are not possible with conventional annotation protocols. Moreover, by means of the arrangement of COG database, it is feasible to uncover the primary functional components of an organism and its potential evolutionary similarities, as well as to systematically identify sets of genes whose presence or absence in the given genome is unexpected and instructive from an evolutionary perspective.