Chapter XLIV
Comparative Genomics and Structure Prediction in Dental Research

Andriani Daskalaki
Max Planck Institute of Molecular Genetics, Germany

Jorge Numata
Free University, Germany

ABSTRACT

Since the completion of the Human Genome Project (HGP) in 2003, the 3.2 billion basepairs which make up the human genome have been sequenced. These sequences contain the plan for the mechanisms controlling the behavior of each cell. The small variations in the DNA sequence that lead to different characteristics, such as facial features, or color, are known as polymorphisms, which also can cause oral diseases. Periodontitis is a chronic infective disease of the gums caused by bacteria present in dental plaque. Severaazl techniques have been developed to regenerate4 periodontal tissues including guided tissue regeneration (GTR), and the use of enamel matrix derivative (EMD). EMD is an extract of enamel matrix and contains amelogenins. This is evidence to show that amelogenins are involved not only in enamel formation, but also in the formation of the periodontal attachment during tooth formation. Comparative sequence analysis is an approach for detecting functional regions in genomic and protein sequences. Motifs, conserved domains, secondary structure characteristics, and functional sites of proteins related to oral health may be compared, revealing the degree of sequence conservation during vertebrate evolution. Secondary and tertiary structures are important in understanding the function of a protein. In a comparative sequence analysis, the most well-known bioinformatics tools that are used are: basic local-alignment search tool (BLAST), multiple-sequence alignment software (ClustalW), and PROSITE, a database of proten families and domains. The PROSITE database consists of biologically significant sites, patterns, and profiles that help to reliably identify to which known protein

family a sequence belongs. Phylogeny Inference Package (PHYLIP) can be used for building phylogenetic trees and a Python-enhanced molecular graphics program (PyMOL) for 3D visualization of proteins.

INTRODUCTION

Dental researchers collaborating with bioinformaticians have achieved advances in oral-health research by actualizing the impact of genetics in oral health. With the help of bioinformatics, a spectrum of questions in dentistry can be addressed.

Comparative genomics approaches are used to identify the functional domains of a protein and suggest similarities for assigning 3-D structures by homology modeling. It is then possible to use classical molecular dynamics simulations to account for the dynamic behavior in protein function. Multiple-sequence analysis of proteins in different species reveals the degree of sequence conservation at the nucleotide and protein levels. Motifs, conserved domains, secondary structure characteristics, and functional sites of proteins related to oral health may be compared, revealing the degree of sequence conservation during vertebrate evolution.

Three-dimensional structure predictions developed by the modeling of conserved domains of proteins support a key role for specific residues in processes like mineralization.

COMPARATIVE GENOMICS

Comparative sequence analysis is an approach for detecting functional regions in genomic and protein sequences. It facilitates the identification of conserved domains, motifs, and distantly related sequences of different organisms, and provides evolutional insights into the underlying biology of organisms (Rubin et al., 2000).

The extracellular matrix of dentin primarily consists of Type I collagen, noncollagenous matrix proteins, and proteoglycans.

Amelogenin and osteocalcin are noncollagenous matrix proteins secreted by the ameloblasts and odontoblasts, respectively. These proteins primarily function in enamel mineralization.

Krishnaraaju et al. (2003) have used bioinformatics tools for multiple-sequence analysis of these proteins in different species.

Phylogenetic analysis using sequence data is used to study sequence relatedness.

STRUCTURE PREDICTION

Secondary and tertiary structures are important in understanding the function of a protein. Frequently, however, such information is not available because neither crystallographic nor nuclear magnetic resonance (NMR) structure determination has been carried out. In this case, structure-prediction methods may help. Homology-based methods are not perfect, and depend on the following:

1. One or more known crystal or NMR 3-D structures
2. Strong sequence similarity of the unknown structures (>25%)

The secondary structure elements may be predicted with good accuracy. However, side-chain rotameters and loop insertions may be far from reality.