Using a mutant database to analyze the phenotypic diversity of RecA protein homologs.

Chris Trezzo, Alberto Roca, Alex McPherson

The RecA protein plays a crucial role in recombinational DNA repair. Over forty years of research on the E. coli RecA protein has produced a wealth of knowledge, but there is still no detailed molecular insight on how the proteins structure relates to its function. Now the genomic era has also added a large number of bacterial RecA orthologs allowing comparative studies. Thus, new computational tools are necessary for organizing, analyzing, and archiving the rich dataset from the recombination literature. We have used a MySQL relational database to catalog 300 bacterial RecA homologs and over 1000 RecA missense mutants thereby enabling the datasets to be searched, sorted, and related to one another.

Regions of the E. coli RecA protein that have been subjected to mutagenesis studies were compared to the conservation information observed in our RecA multiple sequence alignment. A SIFT computational analysis was performed on the alignment to quantitate and predict the deleterious effect of all possible RecA missense mutations. Hot- and cold-spots with respect to RecA mutational flexibility were located. Furthermore, we have examined some RecA mutants where there are conflicting experimental reports regarding the phenotypes measured (neutral vs. deleterious). The experimental and theoretical phenotypic values were mapped onto RecA crystal structures to place the observations in a three-dimensional context. Finally, we have identified unexplored RecA motifs that are suitable targets for saturation mutagenesis.