



We have developed the Structure SNP (StSNP) database which provides extensive opportunities to compare structural nsSNP distributions in human proteins, protein complexes and protein-protein interaction networks. StSNP is a cross-database utilizing information for protein structures from the Protein Data Bank (PDB), human nsSNP information from dbSNP at NCBI, human genes/proteins from NCBI's Locuslink and pathway information from Kyoto Encyclopedia of Genes and Genomes database (KEGG)[2]. StSNP also delivers extensive information about the metabolic pathway(s) that the protein is involved in. Searching and obtaining frequency distribution(s) of a reference to mutated amino acids or creating various queries to statistically analyze nsSNPs is available. An amino acid mutation matrix of StSNP displays the frequency of a particular mutation, and is available to search for a particular mutation within the proteins. The multiple structure-sequence viewer, Friend[3] is used to provide the mapping of nsSNPs onto protein structures, while visualizing their structural locations. This structural visualization helps in understanding the possible effect(s) of the nsSNP(s) on protein structure and function by looking at the consequences of side chain replacement. The nsSNP mapping onto protein structures makes use of "on the fly" modeling using a homology modeling program, provided that the appropriate structural template is available (sequence share at least 50% sequence identity); moreover, a user has the possibility to visualize several known nsSNPs on one modeled structure, which provides comprehensive information about possible variations in protein structure.

The ongoing implementation of new features in the database includes performing molecular dynamics and subsequent analysis of the produced trajectories for structures with and without one or more of the nsSNPs. The molecular dynamics is performed with GROMACS and results are stored in intermediated PDB-files. The files can be downloaded and visualized synchronously in Friend to see the effect of a single amino acid mutation on the behavior of protein structure in solvent. Another future direction involves developing dynamic visualization of the metabolic pathways which will be essential in understanding the effects of nsSNPs on proteins within their network. This will let one search nsSNPs and decide whether the SNP could have an effect on the related pathway's function. StSNP is available for public use at <http://brams.bio.neu.edu/st/stsnp.html>.

## References

1. Sherry ST, Ward M, Sirotkin K: dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res* 1999;9:677-679.
2. Kanehisa M: The KEGG database. *Novartis Found Symp* 2002;247:91-101.
3. Abyzov A, Leslin C, Ilyin VA: An integrated analytical front-end application for bioinformatics, Friend. CSBi Conference at MIT 2003.