A search method for novel protein functional site based on the spatial distribution of disease-associated missense variants

C. Motono, The National Institute of Advanced Industrial Science and Technology

Disease-associated variants tend to cluster around functional sites in the protein structure, and clusters that are not around known functional sites may point to undiscovered functional sites. Those functional sites may help to elucidate new protein functions and to discover novel drug targets. We comprehensively mapped disease-associated variants registered in ClinVar database against human protein structures and searched for 3D variant clusters. About 30% of the 3D variant clusters were not located near ligand binding sites or protein-protein interaction sites, indicating they were candidates for novel functional sites. Undiscovered functional sites are likely to be dynamically formed ligand binding pockets (cryptic binding sites) or allosteric sites. We use mixed solvent molecular dynamics simulations to estimate the functions of the unknown candidate sites. We also report the new methods to detect allosteric effects of mutations using molecular dynamic simulations and residue-residue contact analysis.