Computational analyses on conformational dynamics of two-component regulatory system and NADH:ubiquinone oxidoreductase

SHIBATA Mayu

The structural dynamics is essential to understand protein functions and their mechanisms. The widely used experimental and computational approaches are often employed together to compensate their limitations, however, the combinatorial analyses are costly in time and labor. Therefore, development of a novel approach which can be used for the initial analysis in studying unknown conformational dynamics is expected. Such analysis should be able to uncover a part of dynamics important for the function of any size of proteins at least at residue side chain resolution, even when used by itself. Conformational dynamics consists of two aspects, *i.e.* conformational ensemble and the transitional process among the conformations, and these can be described as the set of physicochemical interactions of amino acid residue pairs. Hence, the conformational dynamics of the proteins should be able to be mostly reconstructed from the collection of residue pair interactions. It is known that the interacting residue pairs can be predicted from a collection of homologous amino acid sequences by Direct Coupling Analysis (DCA). DCA is a computational analysis which quantifies the observed correlation of the amino acid residue types of residue pairs from the amino acid sequences. The residue pairs with strong direct correlations are reported to coincide well with residue contacts in 3D structures. These correlations are considered to have developed by the evolutionary constraints posed on amino acid sequences to maintain functionally and/or structurally important residue interactions, therefore the strong direct correlations are referred as coevolution. To date, a number of works have attempted to utilize DCA to uncover structural dynamics of proteins. In fact, the effectiveness of DCA on uncovering conformational ensemble has been shown by many successful examples. However, to the other aspect of conformational dynamics (the conformational transitional process), the description given by DCA remains fragmental and the continuous transitional process is yet to be explained. These show that information on conformational ensemble is considered to be encoded in homologous amino acid sequences, however, it is unclear whether the continuous conformational transition is encoded in amino acid sequences as well. This thesis examined whether the two aspects of the conformational dynamics can be elucidated from coevolving amino acid residue pairs predicted by DCA in the following two studies. This general introduction is given in Chapter I.

Chapter II describes the study on elucidating novel monomer conformations of DNA-binding response regulator (DBRR) proteins, a group of signaling proteins in two-component system. DBRRs were used in this study due to the simple two-domain architecture typical to DBRRs. The acceptance of the transduced signal at the N-terminal domain activates the entire protein, promoting DNA-binding at the C-terminal domain. Such interdomain cooperation should be facilitated by the residue interactions

between the domains. In fact, such interdomain interactions are reported in inactive monomers. Yet, interdomain residue interactions have not been thoroughly investigated, implying that the novel conformations can exit. Following the conventional approach for conformation prediction by DCA, the novel residue interactions between the two domains were explored. A part of the identified coevolving interdomain residue pairs were suggested to belong to the novel inactive monomeric structures by MD simulation and structural evaluation analyses.

In Chapter III, the paths of residue interactions responsible for the conformational change in NADH:ubiquinone oxidoreductase (also known as Respiratory Complex I: RCI) were studied. RCI is a membrane protein complex forming proton gradient for the ATP synthesis in aerobic respiration by the coupled NADH:ubiquinone redox reaction and proton pumping. RCI was selected for this study due to its biological and medical significance. Recent structural studies suggest that the transition from open to closed conformations is initiated by the binding of quinone (Q), the substrate, under the catalytic cycle of RCI. This transition accompanies the loss of π-bulge in the third helix of ND6 subunit (ND6:H3), which is considered to be crucial for the function. However, the molecular mechanism of this change in ND6 remains unelucidated. Considering that the conformational change at ND6:H3 results from the structural change at the Q-binding site, the paths of residue interactions conveying such change were investigated employing the networks of highly coevolving (evolutionarily selected) residue contacts. This novel approach identified the paths of highly coevolving residue contacts involving 9-11 residues in ND1 and ND6 subunits. These newly identified residue interaction paths were supported by the evolutionary couplings and human mutational data.

Chapter IV summarizes the findings and suggest perspectives. By applying DCA to the homologous amino acid sequences, this thesis successfully elucidated a) conformational ensemble (Chapter II) and b) continuous description of the conformational transition (Chapter III). This thesis is the first to report the latter. These results suggest that the two aspects of the conformational dynamics are encoded in the collection of homologous amino acid sequences in the form of residue contacts through evolution. The insights provided by this work extend the usefulness of DCA-based analysis on elucidating conformational dynamics to include conformational transition. This thesis proposes DCA-based analysis as the initial analysis before performing detailed combinatorial analyses of conventional approaches in studying the unknown conformational dynamics of proteins. Such application will streamline the study on conformational dynamics of proteins, facilitating the further elucidation of the function and functional mechanisms of more proteins. Enriched knowledge basis on protein conformational dynamics will contribute to addressing biologically and medically important challenges, including evaluation/prediction of the mutational effects and more efficient drug design.