Unraveling the evolution of early carbon fixation through dynamic tunnel structural changes in CO-catalyzing enzymes

*Kimiho Omae[1], Takashi Fujishiro[2], Shun'ichi Ishii[3], Shino Suzuki[1]

1. RIKEN Cluster for Pioneering Research, 2. Graduate School of Science and Engineering, Saitama University, 3. X-STAR, Japan Agency for Marine-Earth Science and Technology

Microorganisms use CO as a substrate for their life activities. This is made possible by an enzyme called CO dehydrogenase (CODH), which catalyzes the oxidation-reduction reaction of CO/CO_2 ($CO + H_2O \Leftrightarrow CO_2 + 2H^+ + 2e^-$). Many microorganisms with CODH consume high CO concentrations (˜100%) outside the cell to obtain the reducing power they need for growth. In addition, CODH forms a complex with acetyl-CoA synthase (ACS) to supply CO synthesized from CO_2 as a building block for cell material, and is responsible for the carbon fixation pathway (Wood-Ljungdahl pathway: WLP). The WLP is said to be the most ancient carbon fixation pathway, dating back to the last universal common ancestor (LUCA) of life on Earth, and the origin of CO metabolism via CODH is thought to be equally old.

CODH is a metalloenzyme that coordinates a NiFeS cluster at its active center, and CO/CO_2 gas molecules are transported to the active center via a tunnel that penetrates the surface of the protein. The cysteine residues that coordinate the NiFeS cluster are highly conserved between CODH sequences, but the conservation of other amino acid residues is low. Thus, CODH gene sequences are highly diverse. Structural biology has shown that the structure of the CODH gas tunnel varies depending on the sequence and affects the selectivity of gas molecules and reaction efficiency. Therefore, to understand the evolution of CO metabolism in life on Earth, it is important to elucidate the process of diversification of the tunnel structure of CODH. However, structural biology research on CODH is significantly biased towards model organisms, and the bottleneck of research is that there is only a handful of available three-dimensional structural information.

To fill the gap caused by the lack of structural information on CODH, we focused on AlphaFold (AF), an AI technology that won the Nobel Prize in Chemistry last year. AF is a program that predicts protein structures with high accuracy at the atomic level from amino acid sequences, and it has already provided predicted structures for more than 200 million sequences contained in public protein databases. We first searched for CODH genes in public databases and downloaded 3,388 AF prediction structures covering a wide range of lineages. We automated the alignment of multiple CODH structures and the prediction of tunnels using an exploratory simulation starting from the NiFeS cluster, and predicted gas tunnels from all AF prediction structures. The tunnels were clustered into 15 groups based on the distance between the tunnel ends (the entrance or exit points for gas molecules on the protein surface), and the evolution of tunnel structure was investigated by integrating the phylogenetic tree of CODH and the presence or absence of each cluster. CODH was classified into seven phylogenetic groups, A–G, and was broadly divided into clade A, which includes CODH derived from archaeal WLP, and the other clades (which mainly include CODH derived from bacteria). It was found that many CODHs have around 10 tunnels, and that there are tunnels that are preferred (or eliminated) depending on the CODH clade. In addition, it was found that in clades E and F, the number of tunnels in the phylogenetically newer CODH is halved, and that this is related to the acquisition of ACS (WLP). Interestingly, the gas tunnels leading to the ACS are different in the CODH of clades E and F, and it was found that the tunnel that did not connect to the ACS was lost in each clade. This suggests that an adaptive evolution at the structural level to increase the airtightness of the CODH/ACS and prevent CO leakage to increase the reaction efficiency of WLP occurred independently in clades E and F (i.e. convergent evolution). At the same time, this result supports the idea that CODH/ACS in clade E/F were acquired during evolution, and is significant in that it challenges the conventional model that WLPs have been inherited from LUCA.

Keywords: carbon monoxide dehydrogenase, microbiology, carbon fixation, bioinformatics