Structural Analysis of the *Drosophila Melanogaster*'s GSTome

PETIOT Nicolas;[A] SCHWARTZ Mathieu;[B] DELARUE Patrice;[A] SENET Patrick;[A] NEIERS Fabrice[B] and NICOLAÏ Adrien[A,*]

A) Laboratoire interdisciplinaire Carnot de Bourgogne, UMR 6303 CNRS-Université de Bourgogne, 21078 Dijon, France; B) Molecular Mechanisms of Flavor Perception, *Université de Bourgogne, Inrae, CNRS, 21000 Dijon, France*

adrien.nicolai@u-bourgogne.fr

Glutathione transferase (GST) is a superfamily of ubiquitous enzymes, multigenic in numerous organisms which generally presents an homodimeric structures. For instance, GSTs are involved in numerous functions as chemical detoxification as well as chemoperception in mammals and insects[1]. GSTs catalyze the conjugation of their cofactor, the reduced glutathione (GSH) to xenobiotic electrophilic centers. To achieve this catalytic function, GSTs are made of a ligand-binding site and a GSH binding site per subunit. The GSH-binding site (G-site) is very specific and the hydrophobic substrate binding site (H-site) allows the binding of diverse substrates. In addition, the G-site presents a high conservation among GSTs. The chemical diversification in plant during the evolution was probably an important evolutive driver leading to the GSTs diversification, especially in insects. This encourages the study of insect GSTs to understand how spontaneous mutations/insertions/additions in the sequences modify the stability, selectivity, and the catalytic efficiency of this superfamily of enzymes. The Universal Protein Database (UniProt) registers 36 GST sequences in the fly, *Drosophila melanogaster*'s (*D. mel*), which are distributed in 7 classes and composing its GSTome. GST subunit structures are characterized by a $\beta\alpha\beta\alpha\beta\beta\alpha$ thioredoxin-like fold in its N-terminal part (Domain I) and an α-helical domain in its C-terminal part (Domain II). Domain I is quite conserved among GSTs and contains specific residues critical for GSH binding and catalytic activity whereas Domain II is quite variable both in sequence and structure, and this diversity determines the ample and distinct hydrophobic substrate specificities observed for the different enzymes. Finally, the interface of dimerization between the two subunits also plays a crucial role in the stability and catalytic activity of GSTs.

The goal of the present work is to study the complete GSTome of *D. mel* from its 36 sequences and structures to determine how changes in the amino-acid sequence modifies the structural characteristics of GSTs, particularly in the binding sites and in the interface of dimerization. First, we predicted 3-D atomic structures of each GST using the AlphaFold program developed by DeepMind[2] and compared them with existing X-ray crystallography experimental structures (7 over 36 of them have been resolved). We also characterized their global and local fold using secondary structures predictions and free-energy landscape analysis of internal coordinates such as backbone and side-chain dihedral angles. Second, we used Multiple Sequence Alignment (MSA) technique coupled with AlphaFold predicted 3-D structures to characterize the relationship between sequence and structural fingerprints of GST enzymes,. Finally, we applied Normal Mode Analysis using the Anisotropic Network Model[3] to compute the thermal B-factors of all GSTs of *D. mel*. Particularly, we extracted the flexibility profiles of GST enzymes and identify key residues that are systematically involved in the ligand binding/dimerization processes and thus playing a crucial role in the catalytic function. This methodology will be extended to guide the *in silico* design of synthetic GST enzymes with new/optimal catalytic properties for detoxification applications.

Keywords: Glutathione transferase, *Drosophila melanogaster*, AlphaFold, Multiple Sequence Alignment, Normal Mode Analysis

[1] Schwartz et al. *Biomolecules*, 13(2) :322, **2023**
[2] J. Jumper et al. *Nature*, 596(7873) :583-589, **2021**
[3] L. Yang and R.L. Jernigan. *PNAS*, 106 (30) :12347-12352, **2006**