A012-02-250823

NP³: an integrated platform for natural product-based drug discovery centred on X-ray protein crystallography

Daniela B B Trivella[1*], Joane K. Rustiguel[1], Cristina F. Bazzano[1,2], Rafael de Felicio[1], Luiz G. Alves[1], Raphael Meneghello[1], Marcos G. Cunha[1], Guilherme Telles[2], Andrey Z. Nascimento[1], Ana C. M. Zeri[1]

1 Brazilian Center for Research in Energy and Materials (CNPEM), Campinas-SP, Brazil; [2] University of Campinas, Campinas-SP, Brazil

Email of communicating: daniela.trivella@lnbio.cnpem.br

Keywords: high throughput protein crystallography, drug discovery, natural products, deep learning

Natural products (NP) provide new chemical scaffolds for drug discovery and can probe novel enzyme binding sites and inhibition mechanisms [1]. However, the identification of bioactive natural products and their enzyme binding mechanisms is challenging, sample and time-consuming [2]. We have developed an integrated approach to overcome these gaps, based on high throughput screening, high throughput X-ray protein crystallography, massive mass spectrometry analyses and customized software. This approach is named the NP³ platform (Fig. 1).

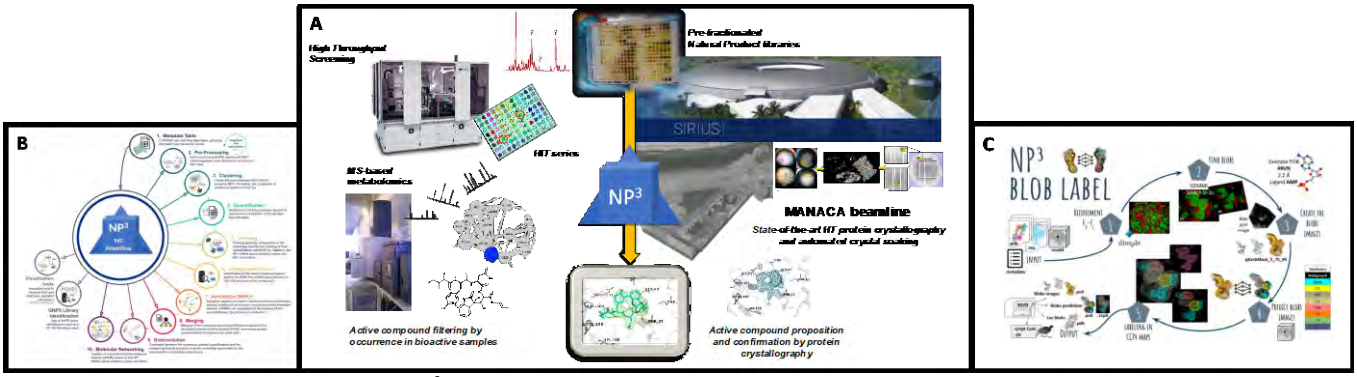

Figure 1. The NP³ platform: experimental routine **(A)** and associated software **(B, C)**.

The NP³ platform starts with the HT screening of pre-fractionated NP libraries, finding bioactive NP samples, which represent mixtures of unknown natural substances. These active samples are then subjected to HT protein crystallography, aiming to capture the bioactive NPs using the crystal itself as the bait. This is done by directly soaking the bioactive NP sample into the protein crystals, following X-ray diffraction data collection, processing and extraction of the residual electron density. The latter, in turn represents the captured NP ligand revealing the active natural product binding site, its mechanism of interaction with the protein, and providing initial clues on its chemical structure. LC-MS/MS-based metabolomics is then employed for filtering candidate *m/z* (compounds) in the unknown mixture. Using an iterative process of crystal electron density and MS/MS spectra interpretation, it is possible to reveal the chemical identity of the bioactive natural product. The experimental data is integrated and mined, using designed computer algorithms for identifying bioactive natural products and their enzyme binding sites, in the very early stages of natural product-based drug discovery. Two software were developed, the *NP³ MS Workflow* (**Fig. 1B**), a software for mass spectrometry data treatment, bioactive NP ranking and chemical structure annotation (available at https://github.com/danielatrivella/NP3_MS_Workflow); and the *NP³ Blob Label* (**Fig. 2C**), a *deep learning application* for unknown ligand segmentation to ligand building in X-ray protein crystallography.

This iterative approach proved successful even when using low resolution protein crystals and active natural products present in trace amounts in complex chemical samples. The process can be performed in miniaturized scales, in which each step is compatible with high throughput techniques. The NP³ platform is empowering natural product drug discovery, as it will be exemplified by target-based drug discovery projects currently running in our pipeline.

1. Bruder, M.; Polo, G.; Trivella, D.B.B. Natural allosteric modulators and their biological targets: Molecular signatures and mechanisms. *Nat. Prod. Rep.* **2020**, *37*.
2. Beutler, J.A. Natural products as a foundation for drug discovery. *Curr. Protoc. Pharmacol.* **2009**, 1–30.

Acknowledgement: Serrrapilheira Institute (Serra-1709-19681/Instituto Serrapilheira)