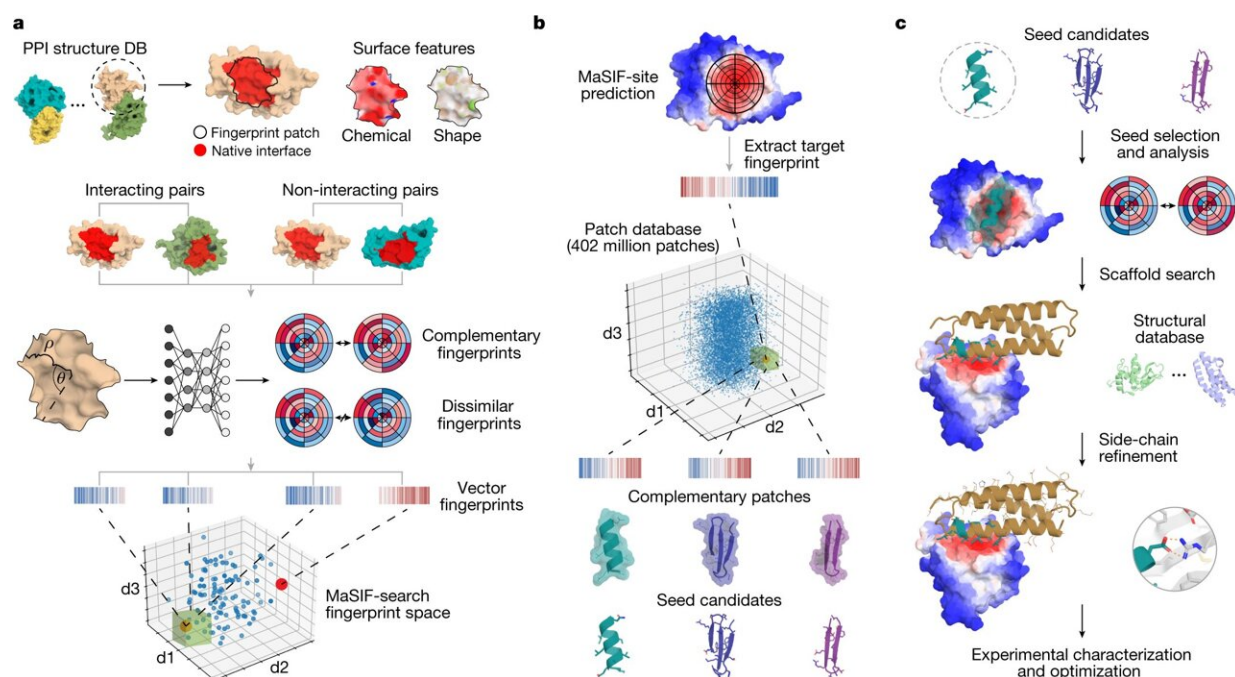# Solving the mystery of protein surface interactions with geometric fingerprints

May 1 2023, by Justin Jackson

Surface-centric design of de novo site-specific protein binders. a, Schematic of fingerprint generation. Protein binding sites are spatially embedded as vector fingerprints. Protein surfaces are decomposed into overlapping radial patches, and a neural network trained on native interacting protein pairs learns to embed the fingerprints such that complementary fingerprints are placed in a similar region of space. We show an illustration for a subsample of the fingerprints projected in a space reduced to three dimensions. The green box highlights a region of complementary fingerprints. b, MaSIF-seed—a method to identify new binding seeds. A target patch is identified by MaSIF-site based on the propensity to form buried interfaces. Using MaSIF-seed, fingerprint complementarity is evaluated between the target patch and all fingerprints in a large database

(around 402 million patches); the pairs of fingerprints are subsequently ranked. The top patches are aligned and rescored to enable a more precise evaluation of the seed candidates. c, Scaffold search, seed grafting and interface redesign. The selected seeds are transferred to protein scaffolds and the rest of the interface is redesigned using Rosetta. The top designs are selected and tested experimentally. Credit: *Nature* (2023). DOI: 10.1038/s41586-023-05993-x

Researchers from the Swiss Institute of Bioinformatics, Lausanne in Switzerland, used a geometric deep-learning tool that generates "fingerprints" of protein surfaces to describe geometric and chemical features critical to protein–protein interactions. In their paper, "De novo design of protein interactions with learned surface fingerprints," published in *Nature*, the team report that their hypothesized "fingerprints" led to the capture of essential aspects of molecular recognition and novel protein interactions. A Research Briefing that summarizes the team's results is published in the same journal issue.

Proteins are the physical machinery of biology, the cogs and gears, springs and valves that allow organic life to function. It is how cells do work, how drugs interact with biological systems and where most diseases interact with biological systems. With a commanding mastery of this machinery, science could cure most pathologies.

Machine learning, along with proteomics, genomic sequencing and molecular biology, have massively accelerated protein research over the past decade to the point where we can predict nearly all existing functional structures, engineer novel structures in any configuration and synthetically manufacture any protein imaginable with one tiny but significant caveat—the binding sites.

The way proteins interact relies on two specific physical "lock-and-key"

interactions based on surface chemistry and structure. There is an outer rim site and a buried site beneath. The buried site does the work of the protein, but to access it, the initiating outer rim signal is required to open the structure to allow access.

As far as proteomics has come in the past decade, the affinity characteristics of proteins have remained elusive. This is partly because proteins are finicky and pH-dependent, with changeable rim surface chemistry and binding sites that are condition and site-specific.

In the current research, the team pursued structural affinity between proteins by essentially ignoring everything but the surface affinity. Setting aside information about the overall structure, function and similar protein interactions to their targets, the team focused the machine learning on the surface interactions of proteins and the geometric and chemical patterns that determine the best chance for two molecules to interact, then designed the appropriate keys.

By computing fingerprints from protein molecular surfaces, the team was able to rapidly and reliably identify complementary surface fragments that can engage a specific target within 402 million candidate surfaces.

Several de novo protein binders were computationally designed to engage four protein targets: SARS-CoV-2 spike, PD-1, PD-L1 and CTLA-4. Several designs were experimentally optimized, whereas others were generated purely in a digital space. The results were highly accurate affinity predictions as the machine-learning-based binders successfully engaged their targets.

The authors state that their framework could "...open possibilities in other important biotechnological fields such as drug design, biosensing or biomaterials in addition to providing a means to study interaction

networks in biological processes at the systems levels."

 More information: Pablo Gainza et al, De novo design of protein interactions with learned surface fingerprints, *Nature* (2023). DOI: 10.1038/s41586-023-05993-x

New protein–protein interactions designed by a computer, *Nature* (2023). DOI: 10.1038/d41586-023-01324-2

Citation: Solving the mystery of protein surface interactions with geometric fingerprints (2023, May 1) retrieved 29 April 2024 from https://phys.org/news/2023-05-mystery-protein-surface-interactions-geometric.html