Enzymes have applications in a wide array of industries due to their ability to catalyze reactions at high rates under mild conditions. Enzymatic catalysis can have many advantages over conventional catalytic processes, namely less energy consumption and fewer unwanted by-products. However, a natural enzyme that can catalyze the desired reaction does not exist for most industrial chemical processes. While there have been successes in the protein design field, including de novo enzyme design for a few different reactions that have no known natural counterpart, the enzymes produced have low activity compared to natural enzymes. Laboratory techniques such as directed evolution have been used to further optimize designed enzymes to improve activity levels, but these methods are costly and very time consuming, limiting their widespread use.

THEMATICS is a computational method that identifies active site residues through their perturbed ionization behavior. Predictions often include not only the residues that directly contact the substrate but also residues farther away. These distant residues may not contact the substrate directly, yet they can still contribute greatly to catalysis, as we have verified through biochemical assays on single-site variants. Most designed enzymes, however, lack these electrostatic properties that we observe essentially universally in the local active regions of natural enzymes. While natural enzymes may have extensive active site networks, designed enzymes are typically less connected, with fewer residues seemingly contributing to catalysis. These differences may be one reason for their low activity, presenting a path towards enzyme optimization by including these properties in design processes. Supported by NSF-MCB-1517290.

- 20 different amino acids
  - hydrophobic, polar, acid/base
- Catalyze essential life reactions
  - up to $10^{26}$ rate enhancement

**Natural versus designed enzymes**

**Retro-Aldolase 95.5-5**
- Designed enzyme (RA95)$^3$
- 5 rounds of directed evolution$^4$
- K83 is catalytic

**Phosphoglucone Isomerase**
- Natural enzyme
- E358 and H389 are catalytic$^5$
- Biochemically verified extended active site$^6$

**RESULTS**

**Computational methods can identify functional residues and interactions**
- Complex interaction networks present in natural enzymes
- Arrangement allows increased mud of multiple residues, rather than just one

**More focus on arrangement of residues needed for enzyme design**
- Prospective principles to design active sites
- Retrospective analysis to understand activity

2. J. Von Durre et al., Bioinformatics 2011, 27, pg 1712.