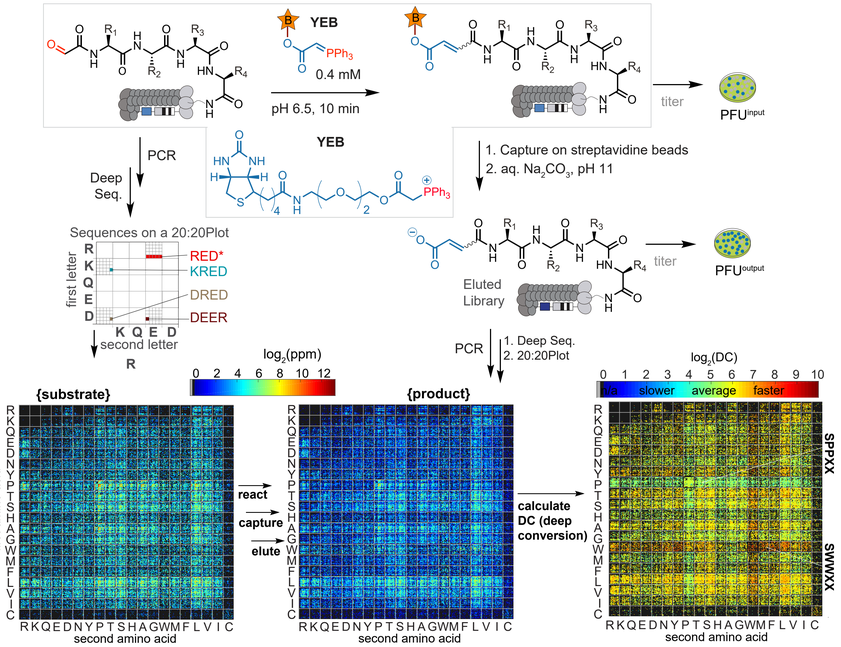
# Learning Structure Activity Relationship (SAR) of the Wittig Reaction from Genetically-Encoded substrates

**Kejia Yan**, Vivian Triana, Kwami Aku-Dominguez, Alex Brown, Ratmir Derda

Department of Chemistry, University of Alberta, Edmonton, AB T6G 2G2, Canada

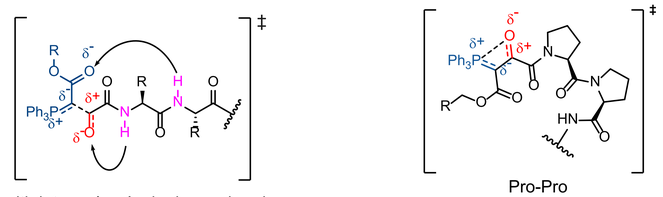
## Introduction

- Reaction rate of Wittig reaction strongly depends on the two N-terminal amino acid.
- Peptides with penultimate Pro-Pro exhibit a significantly decreased reaction rate.




## Plausible hypothesis for SAR

- Hydrogen bonds inside the peptide backbone stabilize the transition state (TS) and accelerate this reaction.

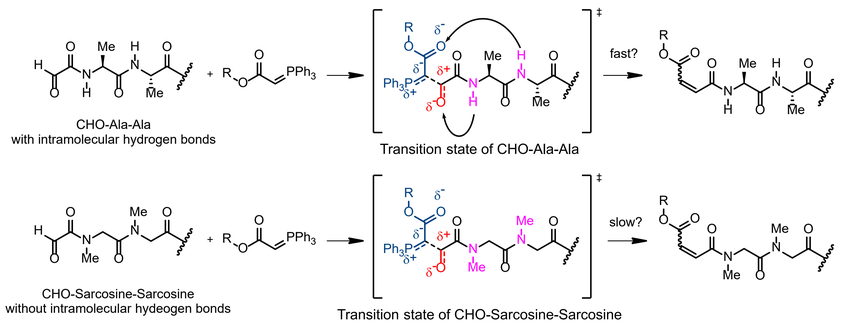


with intramolecular hydrogen bonds accelerate



Pro-Pro with no intramolecular hydrogen bond decelerate

## Experiment design




CHO-Ala-Ala with intramolecular hydrogen bonds

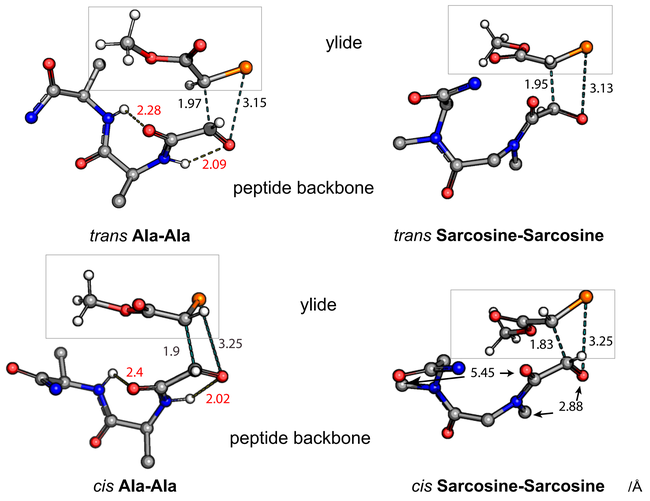
Transition state of CHO-Ala-Ala

CHO-Sarcosine-Sarcosine without intramolecular hydeogen bonds

Transition state of CHO-Sarcosine-Sarcosine

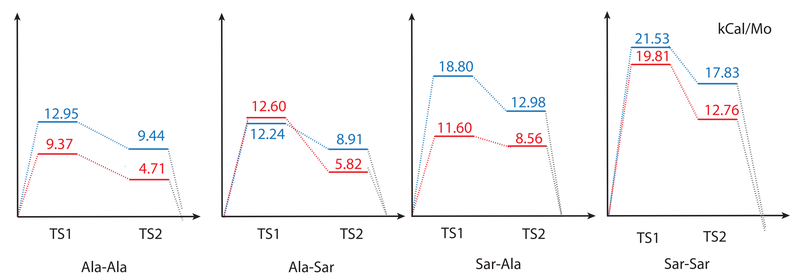
## DFT geometry calculation

- DFT calculation of transition state geometry by gaussian shows in model peptide CHO-Ala-Ala, hydrogen bonds between backbone contribute to stablization of TS.




## DFT Gibbs free energy calculation

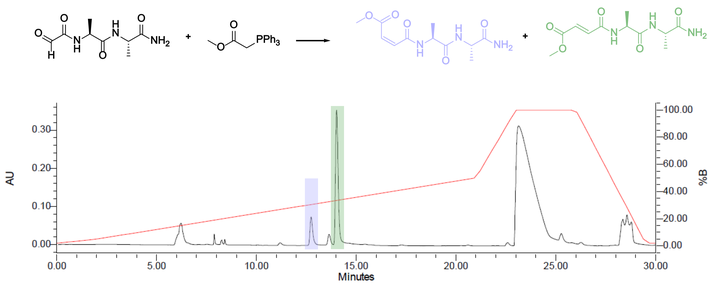
- DFT calculation of gibbs free energy showed that intramolecular hydrogen bonds can stablize transition states.
- *trans* TSs are more favored than *cis* TSs.

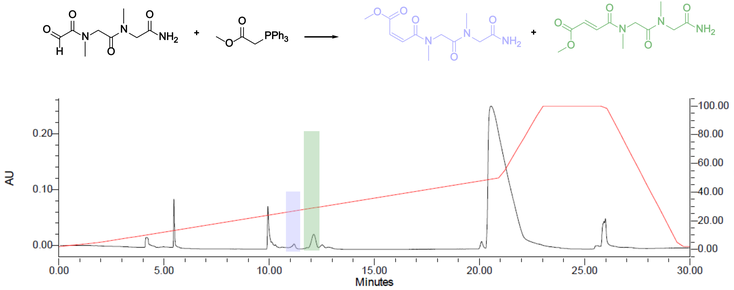



## DFT calculation vs experimental data

- Model peptide with hydrogen bonds showed higher reaction rate.
- *trans* TSs are more favored than *cis* TSs.
- Experimental data fits with DFT calculation.

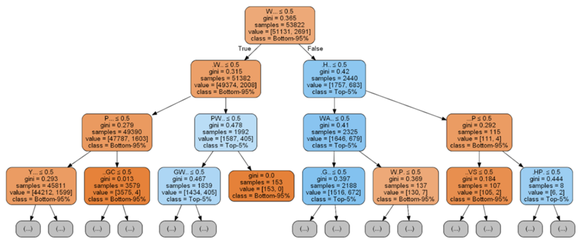
| Peptide Sequence | Ylide | k ($M^{-1}s^{-1}$) *measured by HPLC* |
|---|---|---|
| *HCO*-**AA** | $CH_3OCOCH_2PPh_3$ | 0.093 ± 0.02 |
| *HCO*-**SarSar** | $CH_3OCOCH_2PPh_3$ | 0.021 ± 0.01 |
| *HCO*-**PPAA** | $CH_3OCOCH_2PPh_3$ | 0.035 ± 0.01 |
| *HCO*-**PPAA** | Biotinylated$(PEG)_2$ $OCOCH_2PPh_3$ | 0.024 ± 0.002 |



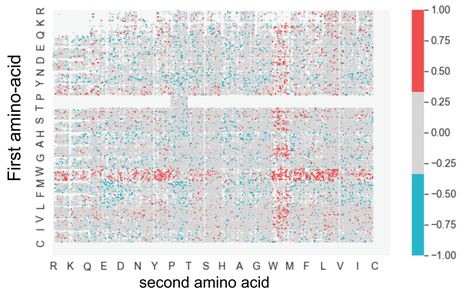

## Apply to machine learning

- XGBoost was used to train the dataset to predict unobserved peptides.



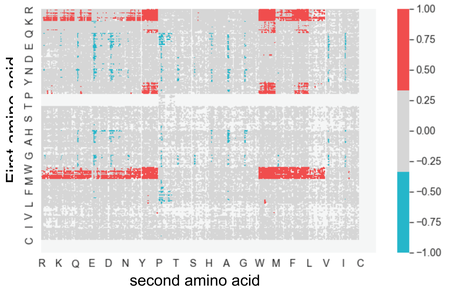
**Original Observed Data (Filtered C and PXXX)**

Observed Data labelled with -1, 0, 1 for bottom 5%, middle 90% and top 5% respectively.




**Predicted Probabilities (Filtered C and PXXX)**

Predicted labels for unobserved sequenes with -1, 0, 1 for bottom 5%, middle 90% and top 5% respectively.




Reference: Triana, V., a Ratmir Derda. "Tandem Wittig/Diels–Alder diversification of genetically encoded peptide libraries." *Organic & biomolecular chemistry* 15.37 (2017): 7869-7877.
Triana, V. 2018, 'Phage display as a Combinatorial Chemistry Platform for Discovery of Chemical Structure-Activity Relationships'', Master thesis, University of Alberta, Edmonton.