# Rational Design of Glycopeptides Enrichment Tool based on Boronic Acid Chemistry and ESI-MS/MS Analysis

**Xu Yang**[1,2], Gordon Chalmers[2], Asif Shajahan[2], Stephanie Archer-Hartmann[2], Christian Heiss[2], Parastoo Azadi[2]

[1]Department of Chemistry, University of Georgia, Athens, GA

[2]Complex Carbohydrate Research Center, University of Georgia, Athens, GA

## Introduction

Many instrumentation and bioinformatics advancements have been made to address the growing demand of mapping GlycoAtlas in a glycoproteomics workstyle. But these fruitful efforts made the bottleneck on glycopeptide enrichment even more pronounced. Lectin or HILIC –based enrichment strategies may be limited by their fundamental specificity scope or bias against aliphatic glycopeptides, which is more substantial in O-GalNAc analysis. The repertoire of boronic acids has been exploited for sensing and enriching carbohydrates for decades, while it has yet been adopted by broader glycoscience community. The selection of local structures of boronic acid by analysts seems empirical, a knowledge gap for screening of molecules for better enrichment practices is left beige if not blank.[1]

## Research Question

1. Could novel binding specificity between boronic acids and glycopeptides be discovered by *in silico* simulations?
2. Will the chemical space of analytical enrichment tool candidates be expanded by calculations out of path dependency?

## Methods

1. The boronic acid candidates were computationally determined by a molecular construction package *Ligand GA*[2]*,* a genetic algorithm-based search algorithm, with binding interaction as a fitness function. An initial population of ligands based on the most easily available molecule, phenylboronic acid, was input in SMILES expression. Following this, very large numbers of extensive docking calculations were used to identify the different binding modes and to quantify the specificity to the glycopeptide versus the peptide.
2. Glycopeptides are substantially smaller and more flexible than their protein counterparts. Ligands of cocrystal entries with glycopeptide binding proteins in PDB. These include 3WV0, 5A2K, 5KDS, 7JTV. The solution NMR structure 1KYJ was selected as current docking target for it bears a cluster of Tn antigens and represents O-GalNAc enrichment targets well.
3. To verify the selectivity of screened boronic acid candidates, docking against deglycosylated 1KYJ, *i.e.,* Ac-STTAV was tested. Considering the working pH of boronic acid enrichment tools, both trigonal and tetrahedral forms of boronic acids were compared.
4. Regarding the enrichment protocol compatibility with Eclipse LC-MS, phenylboronic acid magnetic beads was derivatized and used for Fetuin tryptic digest enrichment per literature protocol.[3]

## Results

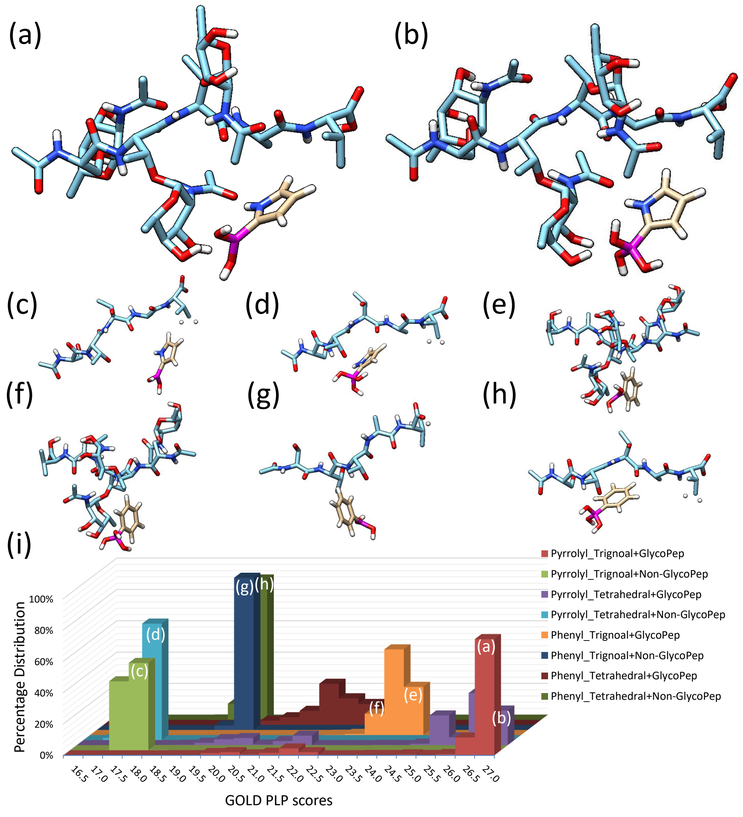
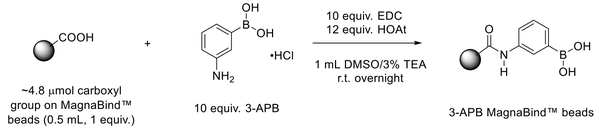



**Figure 1 (a)-(d)** Showing highest PLP scored pre-covalent docking conformations between (2-pyrroyl)boronic acid, the most promising local structure found by Ligand GA iterations, with 1KYJ and deglycosylated 1KYJ under low pH (trigonal form) and high pH (tetrahedral form). **(e)-(h)** Showing the same docking using phenylboronic acid, the most easily available carbohydrate sensor. **(i)** GOLD PLP score percentage distributions of large-scale (up to $5\times10^3$) uniformly randomly sampled dihedral angles. Conformations **(a)-(h)** can be located around the highest scoring bins.



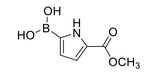

**Scheme 1** Derivatization of magnetic beads using phenylboronic acid

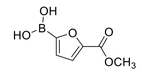
| Sample Treatment | Sequence Coverage (%) | | #Unique Glycopeptides | | Glycosylation Sites | |
|---|---|---|---|---|---|---|
| | N-glycopeptides Search | O-glycopeptides Search | N-glycopeptides Search | O-glycopeptides Search | N-glycopeptides Search | O-glycopeptides Search |
| Fetuin Tryptic Digest | 92.20 | 88.02 | 22 | 13 | N99, N156 | S101, T253, S271, S296, T334, S341 |
| Enrichment with 3-APB beads | 67.40 | 80.20 | 2 | 7 | N99 | S271, T334, S341 |

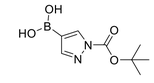
**Table 1** Preliminary result of phenylboronic acid beads enrichment. Lyophilized fetuin tryptic digest was resuspended in 0.5% TEA in DMSO, incubated end-over-end at r.t. for 1 h, then washed by 50% DMSO/50% 100 mM Ammonium Acetate 5 times. Glycopeptides were eluted with water/ACN/TFA 50:49:1 at 37 ℃ for 30 min, then further eluted by 5% FA at 50 ℃ for 5 min twice. Eluates were combined and lyophilized for LC-MS analysis. Glycopeptides were identified by Byonic specific search under a set of strict filters: PEP 2D<0.001, Glycans is not empty, Off-by-X is empty.

## Conclusion

1. The new founded simulated interaction between the secondary amine group on (2-pyrroyl)boronic acid structure and the carbonyl group on O-GalNAc contributed to a three-clawed pre-covalent conformation, which may trigger covalent bonding in higher efficiency than the most easily available phenylboronic acid structure. Moreover, a more drastic selectivity against glycopeptides over non-glycosylated peptides (9.5 ΔPLP score) was observed in 8-atomed (2-pyrroyl)boronic acid than that of 9-atomed phenylboronic acid (4.7 ΔPLP score) *in silico*. Ring geometry may also contribute to the stronger binding.
2. Phenylboronic acid derivatized magnetic beads did not enrich more glycopeptides than native Fetuin tryptic digest, while TEA/TFA used in protocol may suppress signal. Revised protocol must be explored for compatibility. A cocktailed protein standard, such as human plasma might be a better choice for testing enrichment efficiency.
3. Guided by computational results, serval five-member aromatic ring boronic acids were purchased for the development of a new line of enrichment tools.


(5-(methoxycarbonyl)-1*H*-pyrrol-2-yl)boronic acid


(5-(methoxycarbonyl)furan-2-yl)boronic acid


(1-(*tert*-butoxycarbonyl)-1*H*-pyrazol-4-yl)boronic acid

## Acknowledgement

This work was supported by GlycoMIP, a National Science Foundation Materials Innovation Platform funded through Cooperative Agreement DMR-1933525.

## References

1. Riley, N. M.; Bertozzi, C. R.; Pitteri, S. J., A Pragmatic Guide to Enrichment Strategies for Mass Spectrometry–Based Glycoproteomics. *Molecular & Cellular Proteomics* **2021,** *20*, 100029.
2. Chalmers, G., Ligand GA: a genetic algorithm for automated protein inhibitor design, bioRxiv 2021.10.11.463970. Preprint. Software at https://github.com/gordonchalmers/.
3. Xiao, H.; Chen, W.; Smeekens, J. M.; Wu, R., An enrichment method based on synergistic and reversible covalent interactions for large-scale analysis of glycoproteins. *Nature Communications* **2018,** *9* (1), 1692.
4. Molecular graphics and analyses performed with UCSF Chimera. https://www.rbvi.ucsf.edu/chimera/