A Combinatorics Approach to Lipid A Structure Identification

Tao Liang[1]; Michael C. Wilson[1]; Sung Hwan Yoon[1]; Lisa Leung[3]; Robert K. Ernst[2], David R. Goodlett[1]

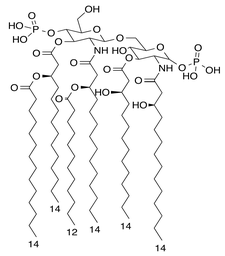
[1] School of Pharmacy and [2] School of Dentistry and [3] School of Medicine at the University of Maryland, Baltimore, MD

Introduction

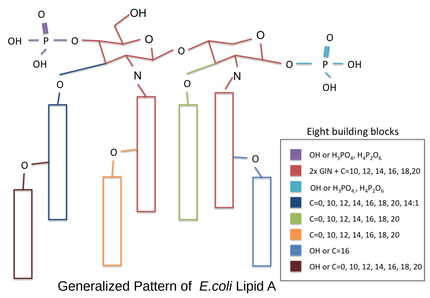
Lipid A, the major component of Gram-negative bacterial outer membrane, is also known as endotoxin and is recognized by host immune system. Recently, we reported on the hierarchical tandem mass spectrometry (HiTMS) algorithm (Ting et al. *JASMS* 2011) for the assignment of lipid A structures, the hydrophobic anchor of lipopolysaccharide. This approach required acquisition of exhaustive tandem MSn data on multiple precursor ions and subsequent fragment ions. However, due to sample availability, it is not an appropriate technique for clinically relevant samples. Therefore, we are developing a combinatorics approach to solve these unique glycolipid structures. This new approach results in a probabilistic match to a structure that requires only tandem mass spectra (MS2).

In this study, theoretical structures for *E. coli* lipid A were computed using a combinatorics algorithm that considered the heterogeneity present in the fatty acids (number of carbon chain length and position variety) and the phosphorylation pattern on the disaccharide backbone. Precursor ion masses detected from a lipid A structure were compared against the MS1 theoretical database we constructed and the top n selected based on a match to the available molecular masses. A hypothetical structure is further elucidated by matching the tandem mass spectra against a secondary MS/MS database consisting of all the possible fragmentation patterns based on the combinatorics algorithm.

MS1 Theoretical Database Construction



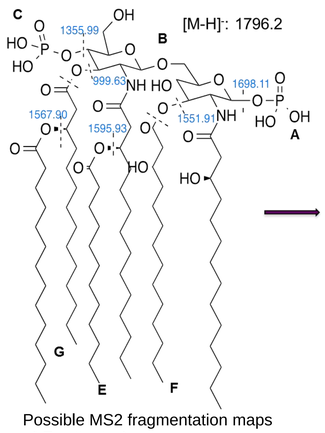
W.T Structure of *E.coli* lipid A



Generalized Pattern of *E.coli* Lipid A

Based on manual interpretation of lipid A precursor ion (MS^1) data and fragmentation rules (Ting et al. *JASMS* 2011), eight components (fatty acid, phosphate pattern, disaccharide et al.) of potential theoretical masses derived from the *E. coli* lipid A structure were calculated. Using our combinatorial algorithm, each component is considered as a module to construct a hypothetical lipid A structure.

MS2 Theoretical Database Construction



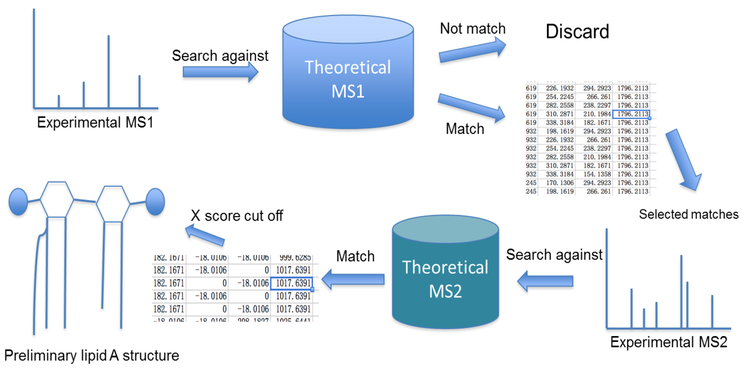
Possible MS2 fragmentation maps of *E.coli* lipid A

A	B	C	D	E	F	G	H
One	Disaccharide	FourPrime	TwoB	TwoPrimeB	Three	3 And 3'B	theoretical data
79.9663	792.5347	79.9663	0	182.1671	208.1827	-18.0106	1323.7987
79.9663	792.5347	79.9663	0	182.1671	226.1932	-18.0106	1341.8092
79.9663	792.5347	79.9663	0	182.1671	0	208.1827	1341.8093
79.9663	792.5347	79.9663	0	182.1671	208.1827	0	1341.8093
79.9663	792.5347	79.9663	0	-18.0106	208.1827	208.1827	1349.8143
79.9663	792.5347	79.9663	0	-18.0106	-18.0106	436.3916	1351.8299
-18.0106	792.5347	-18.0106	0	182.1671	208.1827	208.1827	1354.0382
-18.0106	792.5347	-18.0106	0	182.1671	-18.0106	436.3916	1356.0538
79.9663	792.5347	79.9663	0	182.1671	226.1932	0	1359.8198
79.9663	792.5347	79.9663	0	-18.0106	226.1932	208.1827	1367.8248
79.9663	792.5347	79.9663	0	-18.0106	0	436.3916	1369.8405
-18.0106	792.5347	-18.0106	0	182.1671	226.1932	208.1827	1372.0487
-18.0106	792.5347	-18.0106	0	182.1671	0	436.3916	1374.0644
-18.0106	792.5347	-18.0106	0	-18.0106	208.1827	436.3916	1382.0694
-18.0106	792.5347	-18.0106	0	-18.0106	226.1932	436.3916	1400.0799
79.9663	792.5347	-18.0106	0	182.1671	208.1827	208.1827	1452.0151
-18.0106	792.5347	79.9663	0	182.1671	208.1827	208.1827	1452.0151
79.9663	792.5347	-18.0106	0	182.1671	-18.0106	436.3916	1454.0307
-18.0106	792.5347	79.9663	0	182.1671	-18.0106	436.3916	1454.0307
79.9663	792.5347	-18.0106	0	182.1671	226.1932	208.1827	1470.0256

Searching results of *E.coli* lipid A MS2 database

For MS2, possible fragment patterns determined from literatures reported tandem mass spectrometry data (John P. O'Brien, et. al, *Ana. Chem*. 2014, Chang-soo Lee et al., *JASMS*, 2004). Based on these, *E.Coli* lipid A fragmented map was drawn and all possible fragment ion masses were calculated.

Proposed Flowchart of Combinatorics Algorithm



All possible component configurations were calculated to generate theoretical *E. coli* lipid A MS1 data (~300,000 possible combinations) and incorporated into the program implemented in Python 2.7.6. Acquired MS1/MS2 experimental data were manually verified by searching against lipid A MS1/MS2 theoretical database.

Conclusions

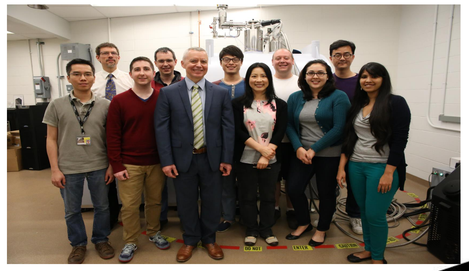
- All possible exact masses for precursor ion were calculated based on the generalized *E.coli* lipid A structure resulting in a total of 296,352 possible combinations and incorporated into a MS1 lipid A database.
- All possible exact masses for fragment ions were calculated based on the proposed MS2 fragmentation map and manually verified by using previously published diagnostic fragment ions. Of these 12 featured fragment ions, all were matched in MS2 theoretical database which contained 128 possible ions for m/z of 1796 $[M-H]^-$ precursor ion.
- Preliminary results of our combinatorics algorithm for Lipid A structure identification are promising. While the initial work was focused on elucidating the lipid A structure of *E.coli*, we are working on testing the approach against a variety of bacterial species lipid A structures that show heterogeneity in both the length and number of fatty acids as well as modification of the phosphate group attached to the diglucosamine backbone.

References

1. Ting, Ying S., et al. *J. Am. Soc. Mass Spectrum* **22** (2011): 856-866.
2. Chang-soo Lee et al., *J. Am. Soc . Mass Spectrum., **39*** (2004): 514-525.
3. O'Brien, John Patrick, et al. *Anal. Chem.* **86** (2014) : 2138-2145
4. Needham, Brittany D., and M. Stephen Trent. *Nat. Rev. Micro*. **11** (2013): 467-481.

Acknowledgements

- This work was supported in part by the University of Maryland Baltimore, School of Pharmacy Mass Spectrometry Center (SOP1841-IQB2014).



Goodlett Laboratory