A Cartesian Product Approach To Lipid A Structure Identification

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

[1] University of Maryland Baltimore, School of Pharmacy, Baltimore, MD; [2] University of Maryland Baltimore, School of Dentistry, Baltimore, MD

Goodlett Laboratory

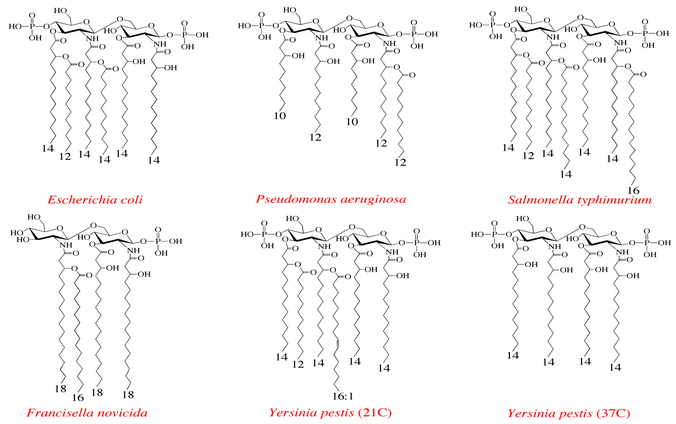
INTRODUCTION:

Building upon our previous work, a combinatorics approach to lipid A structure identification [1,2], we set out to produce a complete theoretical MS^1 database which will allow us to identify the correct lipid A structure without the need for MS^n ($n \geq 3$) data. Previously, only *E. coli* lipid A was considered in the creation of the database. This was primarily due to the sheer size of the database that would be generated.

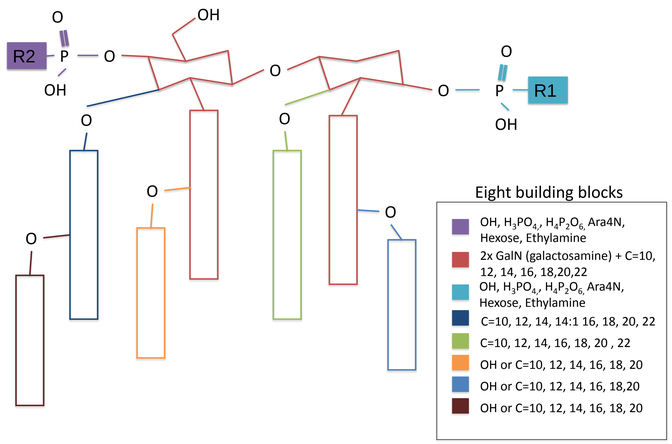
In this study, a subsets approach was taken to identify all possible masses for each component of lipid A. A Cartesian product algorithm was then used to compute the theoretical arrangements of lipid A based on our criteria of even chain fatty acids. Data were then imported into a MySQL database for organization and searching. The database will be cleaned of known impossible arrangements for faster search time and fewer misleading results.

Comprehensive MS^1 Theoretical Database:

Lipid A Structural Variations

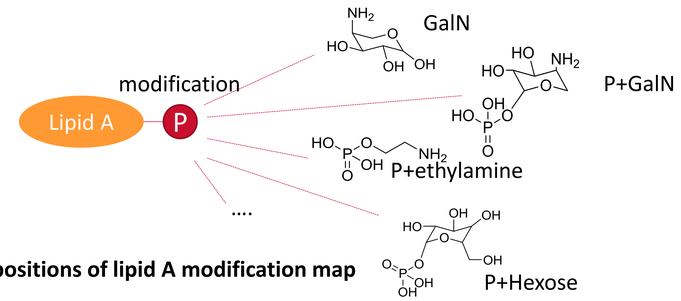


All the Theoretical Lipid A Components



Structural Modification on Phosphate Groups

1, 4' Position	Formula	Exact mass	Database mass	Database Formula
N/A	N/A	0	0	N/A
Phosphoric acid(P)	H_3PO_4	97.9769	79.9663	P-H_2O
P+P	$H_4P_2O_6$	161.9483	143.9377	P+P-H_2O
GalN (hexosamine)	$C_6H_{13}NO_5$	179.0794	161.0688	GalN-H_2O
P+GalN	$C_6H_{14}NO_8P$	259.0457	241.0351	P+GalN-H_2O
P+P+GalN(hexosamine)	$C_6H_{15}NO_{11}P_2$	339.012	321.0014	P+P+GalN-H_2O
Hexose (i.e.: glucose)	$C_6H_{12}O_6$	180.0634	162.0528	Hex-H_2O
P+Hexose	$C_6H_{13}O_9P$	260.0297	242.0191	P+Hex-H_2O
P+P+Hexose	$C_6H_{14}O_{12}P_2$	339.9960	321.9854	P+P+Hex-H_2O
Ara4N (arabinose)	$C_5H_{11}NO_4$	149.0688	131.0582	Ara4N -H_2O
P+Ara4N	$C_5H_{12}NO_7P$	229.0351	211.0245	P+Ara4N-H_2O
P+P+Ara4N	$C_5H_{13}NO_{10}P_2$	309.0015	290.9909	P+P+Ara4N-H_2O
P+ethylamine	$C_2H_8NO_4P$	141.0191	123.0085	PEtN-H_2O
P+P+ethylamine	$C_2H_9NO_7P_2$	220.9854	202.9748	P+PEtN-H_2O



1,4' positions of lipid A modification map

Cartesian Product Algorithm

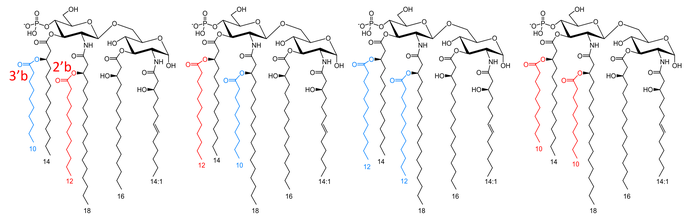
The Cartesian product is equivalent to nesting for loops. For every outer loop advance, the inner loop completes an entire cycle. If the input is sorted, the output will also be sorted. Unlike the previously used Combinatorics approach which produces sorted order without repeats, the Cartesian product approach includes repeated elements of both positional and value types, which generates all possible lipid A combinations.

Comparison of Combinatoric Generators [3]

Iterator	Results
Product('ABCD', repeat=2)	AA AB AC AD BA BB BC BD CA CB CC CD DA DB DC DD
Permutations('ABCD', 2)	AB AC AD BA BC BD CA CB CD DA DB DC
Combinations('ABCD', 2)	AB AC AD BC BD CD
Combinations with replacement ('ABCD', 2)	AA AB AC AD BB BC BD CC CD DD

Comparison of Combinatoric Generators for lipid A Structure Generation

Different combinatoric generators have been used to generate potential lipid A structures. As the example shows below, using Cartesian product function can generate all possible lipid A structures. *A* and *B* stand for a fatty acid that has 10 and 12 carbons respectively at either 3'b or 2'b position. '√' means generated by an iterator function while 'x' means unable to generate a lipid A structure.



Iterator \ Possible lipid A structure	*AB*	*BA*	*BB*	*AA*
Product	√	√	√	√
Permutations	√	√	X	X
Combinations	√	X	X	X
Combinations with replacement	√	X	√	√

Conclusions & Future Directions

- The Cartesian product algorithm produced nearly 2.2 billion theoretical lipid A molecular masses, even though >99.9% of them are not valid with no biologic meaning but could be used as a good decoy database to produce a false positive hit rate.
- Next, 'filter rules' will be incorporated into the program to remove known impossible arrangements for faster searching and fewer false positive hits.
- For MS^2 theoretical database, a machine learning approach[4] will be used to simulate the CID fragmentation process of lipid A molecules in an ion trap instrument. This algorithm does not rely on the chemical reaction equations and fragmentation rules from experimental results.

References:

1.Ting, Ying S., et al. *J. Am. Soc. Mass Spectrum* 22 (**2011**): 856-866.
2.Tao Liang et al. *62nd ASMS poster, **2014.***
3. docs.python.org/2/library/itertools
4. Kangas, Lars J., et al. *Bioinformatics* 28.13 (**2012**): 1705-1713.

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

Posters from Goodlett Lab and Collaborators

MP073 UltrAWN-PTR-MS: Ultrasonic Acoustic Wave Nebulization coupled with Proton-Transfer-Reaction Mass Spectrometr, Lucas Maerk
MP076 Native MS using SAWN, a Novel Ionization Source for Waters SYNAPT G2, Gloria Yen
MP155 Bacterial Glycolipids Characterized on an IMS-QExactive, Yue Huang
MP180 Absorption Mode Analysis of FT-ICR Imaging Data Improves Peak Resolution in a Bordetella pertussis Infection Model, Alison Scott
MP217 Characterization of a Monoclonal Antibody (mAb) using Multiple Fragmentation Techniques and Novel FT Data Processing Software, Bao Tran
MP560 The Associations between Enterovirus Infections and Type 1 Diabetes, Niina Lietzen
MP646 Characterization of Semi-Synthetic Motor Oil using FT-ICR, Sung Hwan Yoon
TP018 Defining Limit of Detection of Mini Surface Acoustic Wave Nebulization Chip by Using Different Types of Mass Spectrometer, Tao Liang
TP122 Absorption Mode Gets Even Better with its Svelte New Curves, David Kilgour

WP233 Direct Beverage Analysis by SAWN MS, David Goodlett
WP268 A Cartesian Product Approach To Lipid A Structure Identification, Lisa Leung
WP450 Ultrasonic Acoustic Wave Nebulization-Mass Spectrometry (UltrAWN-MS) for Unconventional Explosives Characterization, Ben Oyler
WP475 Bridging the Gap between Ion Mobility Spectrometry and an Orbitrap, Mike Belov
WP595 Use of Native Mass Spectrometry for Quantification of Protein Complex, Wenjing Li

UNIVERSITY of MARYLAND SCHOOL OF PHARMACY