

Supplementary Information

Direct Glycan Analysis of Biological Samples and Intact Glycoproteins by Integrating Machine Learning Driven- Surface-Enhanced Raman Scattering (SERS) and Boronic Acid Array

Qiang Hu¹, Hung-Jen Wu^{1,*}

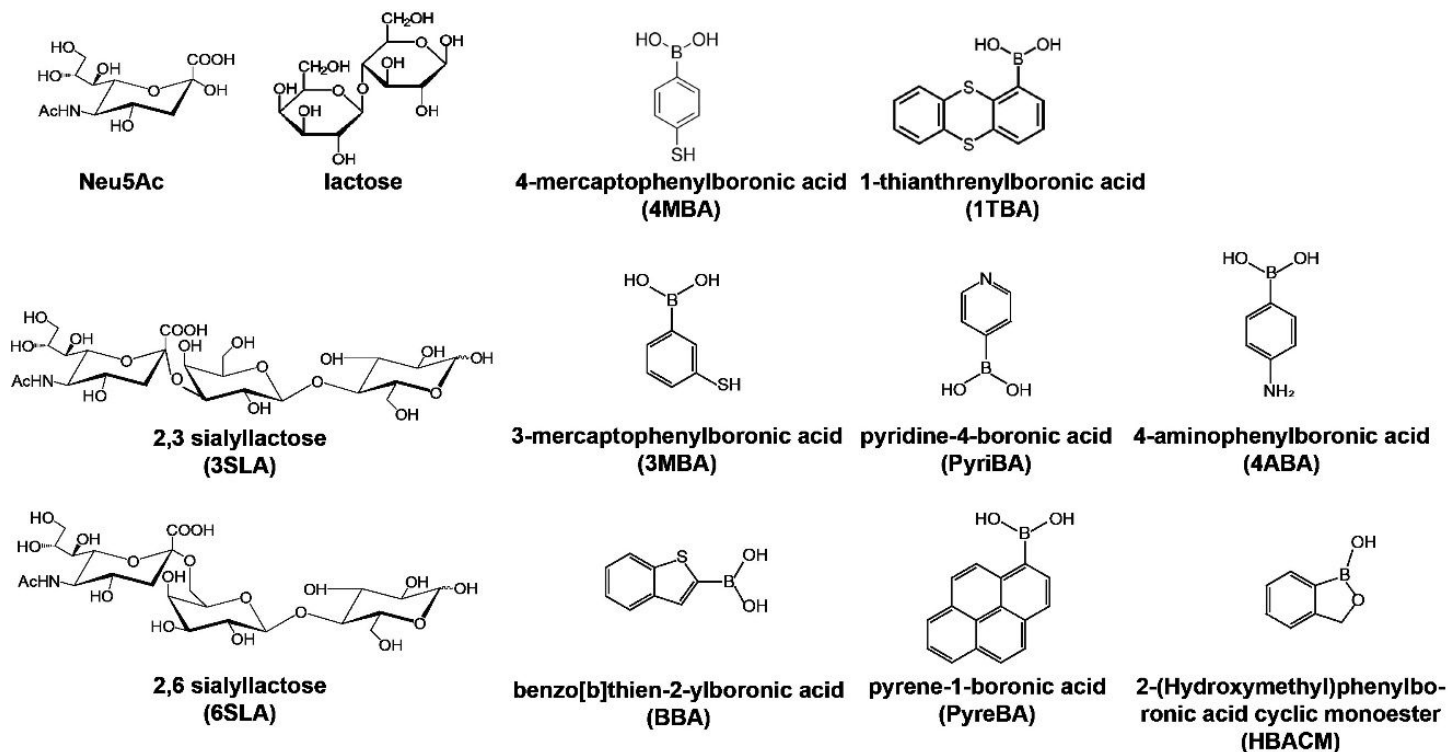
¹ The Artie McFerrin Department of Chemical Engineering, Texas A&M University, TX 77843, USA

* Email: hjwu@tamu.edu

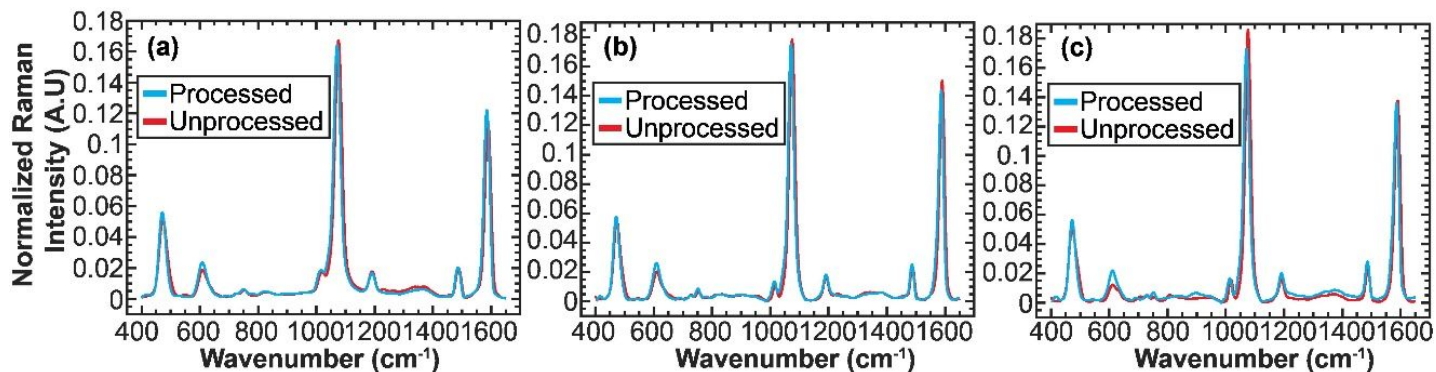
Table of Contents

SI Figure 1. Chemical structures for chemicals (boronic acids and glycans) used in the paper.	4
SI Figure 2 Average normalized SERS spectra (n=200) of the unprocessed cow milk sample and the purified milk oligosaccharides (A), unprocessed goat milk sample and purified oligosaccharides (B), and unprocessed soy milk sample and purified oligosaccharides (C) on 4MBA functionalized substrates. The milk oligosaccharides were extracted from milk using Folch method. The spectral difference between processed and unprocessed samples is minimal.	5
SI Figure 3. The difference between average normalized SERS spectra (n=200) of the unprocessed cow milk sample and the purified milk oligosaccharides (A), unprocessed goat milk sample and purified oligosaccharides (B), and unprocessed soy milk sample and purified oligosaccharides (C) on 4MBA functionalized substrates. The milk oligosaccharides were extracted from milk using Folch method. The spectral difference between processed and unprocessed samples is minimal, and the difference between different milk types were obvious.	5
SI Figure 4. Averaged normalized spectra of cow milk, goat milk, soy milk, oat milk, and almond milk on 4MBA functionalized substrates (n=200 for each milk type).	6
SI Figure 5. Confusion matrix of whole milk classification on 4MBA functionalized substrates (99.9% accuracy).....	6
SI Figure 6. Average normalized spectra of cow milk, goat milk, soy milk, oat milk, and almond milk on 1TBA functionalized substrates (n=200 for each milk type).....	7
SI Figure 7. Confusion matrix of whole milk classification on 1TBA functionalized substrates (99.4% accuracy).....	7
SI Figure 8. Average normalized spectra of cow milk, goat milk, soy milk, oat milk, and almond milk on 3MBA functionalized substrates (n=200 for each milk type).	8
SI Figure 9. Confusion matrix of whole milk classification on 3MBA functionalized substrates (99.0% accuracy).....	8
SI Figure 10. Average normalized spectra of cow milk, goat milk, soy milk, oat milk, and almond milk on 4ABA functionalized substrates (n=200 for each milk type).	9
SI Figure 11. Confusion matrix of whole milk classification on 4ABA functionalized substrates (98.9% accuracy).....	9
SI Figure 12. Average normalized spectra of cow milk, goat milk, soy milk, oat milk, and almond milk on PyriBA functionalized substrates (n=200 for each milk type).	10
SI Figure 13. Confusion matrix of whole milk classification on PyriBA functionalized substrates (99.2% accuracy).....	10
SI Figure 14. Confusion matrix on collective spectra (4MBA, 1TBA, 3MBA, 4ABA, PyriBA) of whole milk classification (100% accuracy).....	11
SI Figure 15. Average normalized spectra (n= 200) of fetuin and asialofetuin on 1TBA, 3MBA, 4ABA and PyriBA functionalized substrates.....	12

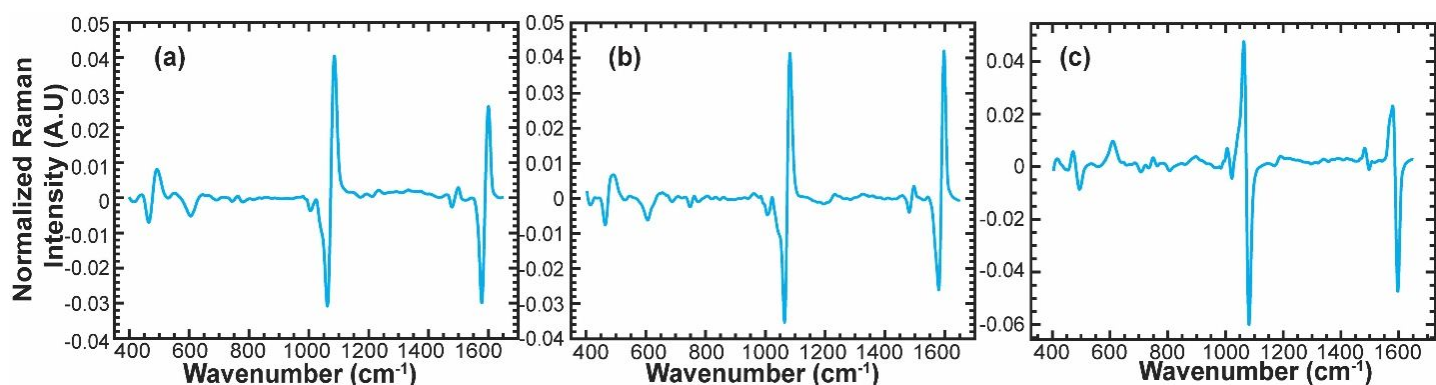
SI Figure 16. Predicted fetuin fraction vs. actual fetuin fraction (A) and residual vs. actual fetuin fraction (B) for 4MBA functionalized substrate.	13
SI Figure 17. Predicted fetuin fraction vs. actual fetuin fraction (A) and residual vs. actual fetuin fraction (B) for 1TBA functionalized substrate.	13
SI Figure 18. Predicted fetuin fraction vs. actual fetuin fraction (A) and residual vs. actual fetuin fraction (B) for 3MBA functionalized substrate.	14
SI Figure 19. Predicted fetuin fraction vs. actual fetuin fraction (A) and residual vs. actual fetuin fraction (B) for 4ABA functionalized substrate.	14
SI Figure 20. Predicted fetuin fraction vs. actual fetuin fraction (A) and residual vs. actual fetuin fraction (B) for PyriBA functionalized substrates.	14
SI Figure 21. Confusion matrix for sialic acid linkage case (SA+Lac, 3SLA and 6SLA) using 4MBA (99.5% accuracy) and 1TBA functionalized substrates (93.8% accuracy).	15
SI Figure 22. Average normalized spectra (n = 200) for 2,3 sialyllactose (3SLA), 2,6 sialyllactose (6SLA), and sialic acid with lactose (SA+Lac) on 4MBA and 1TBA functionalized substrates.	16
SI Figure 23. Confusion matrix for sialic acid linkage case (SA+Lac, 3SLA and 6SLA) using 3MBA (100% accuracy) 4ABA (100% accuracy) and PyriBA functionalized substrates (100% accuracy).....	17
SI Figure 24. Average normalized spectra (n = 200) for 2,3 sialyllactose (3SLA), 2,6 sialyllactose (6SLA), and sialic acid with lactose (SA+Lac) on 3MBA, 4ABA, and PyriBA functionalized substrates.....	18
SI Figure 25. Confusion matrix for sialic acid linkage case (SA+Lac, 3SLA and 6SLA) using PyreBA (96.5% accuracy) HBACM (99.5% accuracy) and BBA functionalized substrates (100% accuracy).	19
SI Figure 26. Average normalized spectra (n = 200) for 2,3 sialyllactose (3SLA), 2,6 sialyllactose (6SLA), and sialic acid with lactose (SA+Lac) on 3MBA, 4ABA, and PyriBA functionalized substrates.....	20



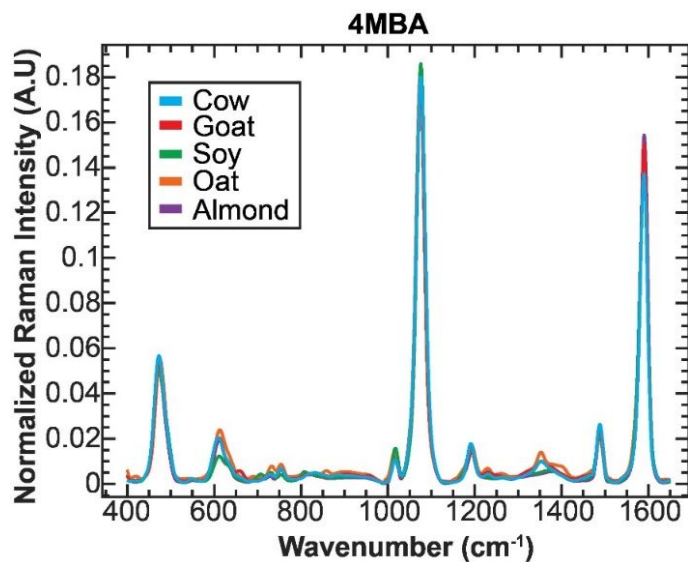
SI Figure 1. Chemical structures for chemicals (boronic acids and glycans) used in the paper.



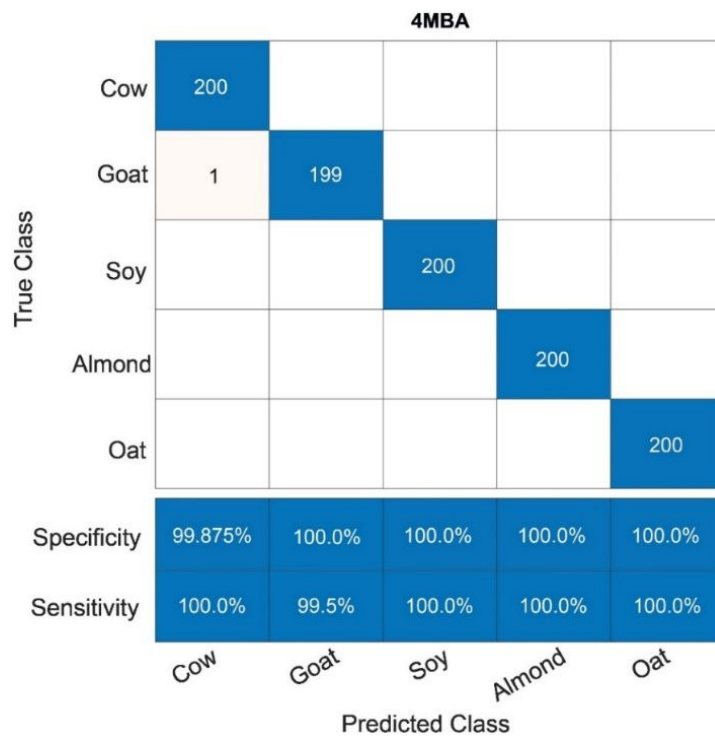
SI Figure 2 Average normalized SERS spectra (n=200) of the unprocessed cow milk sample and the purified milk oligosaccharides (A), unprocessed goat milk sample and purified oligosaccharides (B), and unprocessed soy milk sample and purified oligosaccharides (C) on 4MBA functionalized substrates. The milk oligosaccharides were extracted from milk using Folch method. The spectral difference between processed and unprocessed samples is minimal.



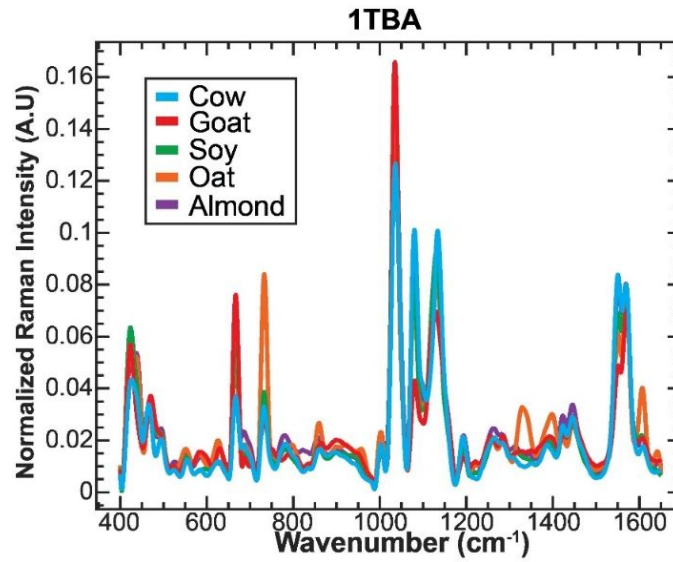
SI Figure 3. The difference between average normalized SERS spectra (n=200) of the unprocessed cow milk sample and the purified milk oligosaccharides (A), unprocessed goat milk sample and purified oligosaccharides (B), and unprocessed soy milk sample and purified oligosaccharides (C) on 4MBA functionalized substrates. The milk oligosaccharides were extracted from milk using Folch method. The spectral difference between processed and unprocessed samples is minimal, and the difference between different milk types were obvious.



SI Figure 4. Averaged normalized spectra of cow milk, goat milk, soy milk, oat milk, and almond milk on 4MBA functionalized substrates (n=200 for each milk type).



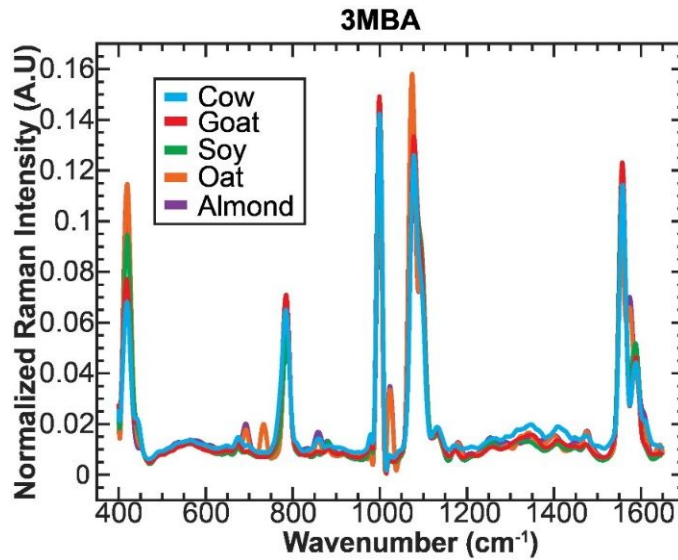
SI Figure 5. Confusion matrix of whole milk classification on 4MBA functionalized substrates (99.9% accuracy).



SI Figure 6. Average normalized spectra of cow milk, goat milk, soy milk, oat milk, and almond milk on 1TBA functionalized substrates (n=200 for each milk type).

		1TBA				
True Class	Cow	198	1	1		
	Goat		200			
	Soy			199	1	
	Almond			1	199	
	Oat		1	1		198
	Specificity	100.0%	99.75%	99.62%	99.87%	100.0%
	Sensitivity	99.0%	100.0%	99.5%	99.5%	99.0%
		Cow	Goat	Soy	Almond	Oat
		Predicted Class				

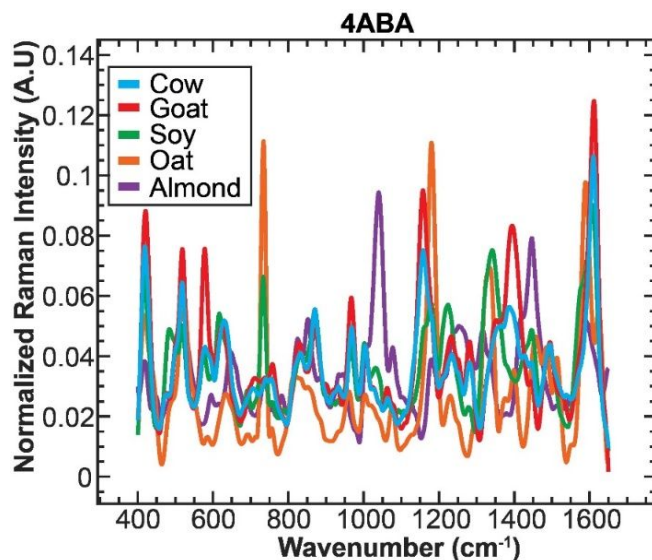
SI Figure 7. Confusion matrix of whole milk classification on 1TBA functionalized substrates (99.4% accuracy).



SI Figure 8. Average normalized spectra of cow milk, goat milk, soy milk, oat milk, and almond milk on 3MBA functionalized substrates (n=200 for each milk type).

		3MBA				
		Cow	Goat	Soy	Almond	Oat
True Class	Cow	198	2			
	Goat	4	196			
	Soy			200		
	Almond				199	1
	Oat				3	197
Specificity		99.499%	99.749%	100.0%	99.62%	99.87%
Sensitivity		99.0%	98.0%	100.0%	99.5%	98.5%
		Predicted Class				
		Cow	Goat	Soy	Almond	Oat

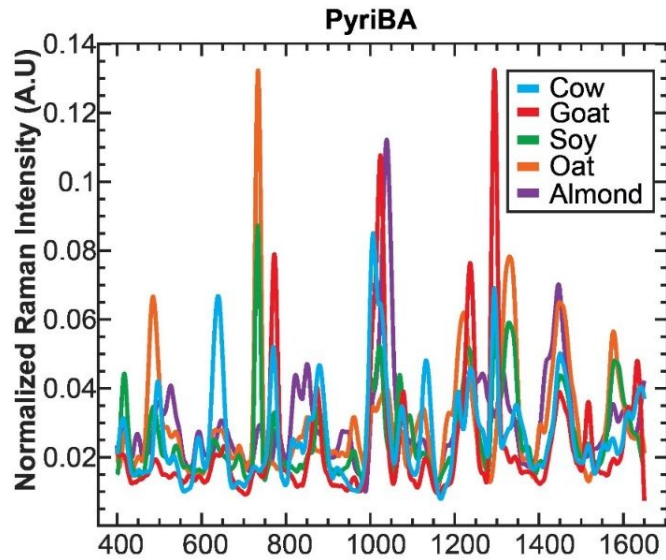
SI Figure 9. Confusion matrix of whole milk classification on 3MBA functionalized substrates (99.0% accuracy).



SI Figure 10. Average normalized spectra of cow milk, goat milk, soy milk, oat milk, and almond milk on 4ABA functionalized substrates (n=200 for each milk type).

		4ABA				
True Class	Cow	195	4	1		
	Goat	1	199			
	Soy	5		195		
	Almond				200	
	Oat					200
	Specificity	99.25%	99.49%	99.87%	100.0%	100.0%
	Sensitivity	97.5%	99.5%	97.5%	100.0%	100.0%
		Cow	Goat	SOY	Almond	Oat
		Predicted Class				

SI Figure 11. Confusion matrix of whole milk classification on 4ABA functionalized substrates (98.9% accuracy).



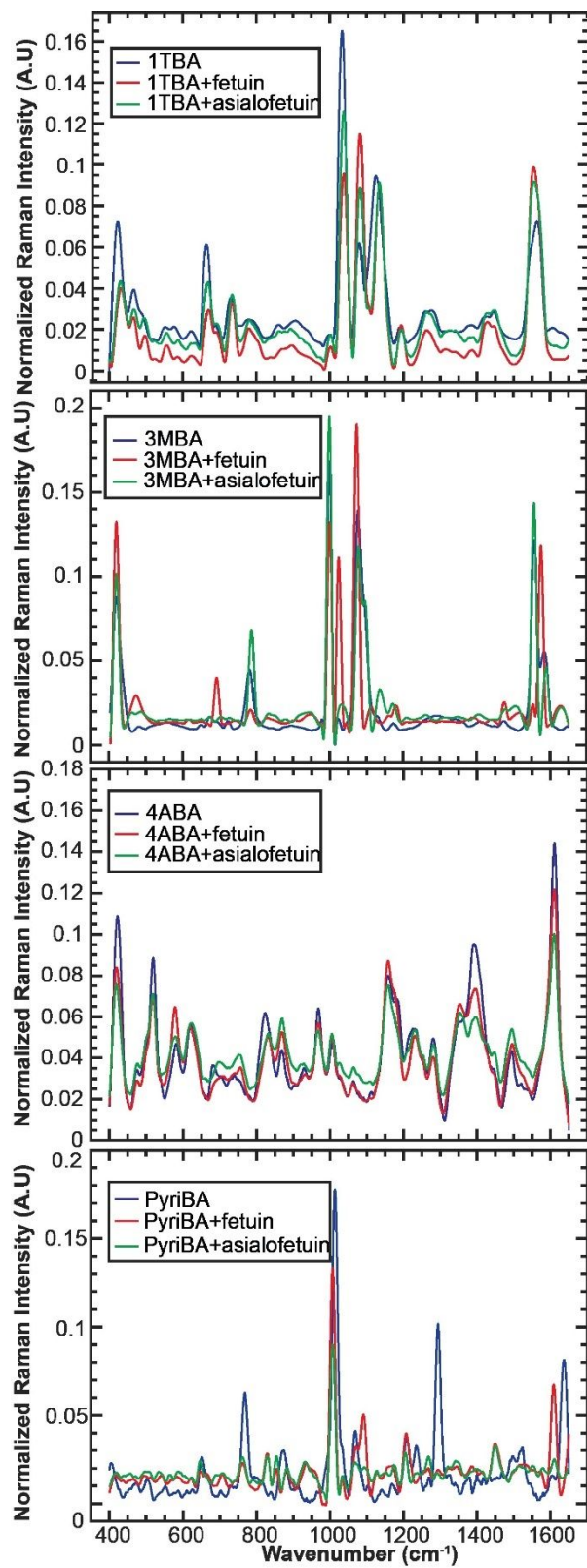
SI Figure 12. Average normalized spectra of cow milk, goat milk, soy milk, oat milk, and almond milk on PyriBA functionalized substrates (n=200 for each milk type).

		PyriBA				
		Cow	Goat	Soy	Almond	Oat
True Class	Cow	196	3	1		
	Goat	2	197	1		
	Soy			200		
	Almond			1	199	
	Oat					200
Specificity		99.749%	99.624%	99.625%	100.0%	100.0%
Sensitivity		98.0%	98.5%	100.0%	99.5%	100.0%
		Cow	Goat	Soy	Almond	Oat
		Predicted Class				

SI Figure 13. Confusion matrix of whole milk classification on PyriBA functionalized substrates (99.2% accuracy).

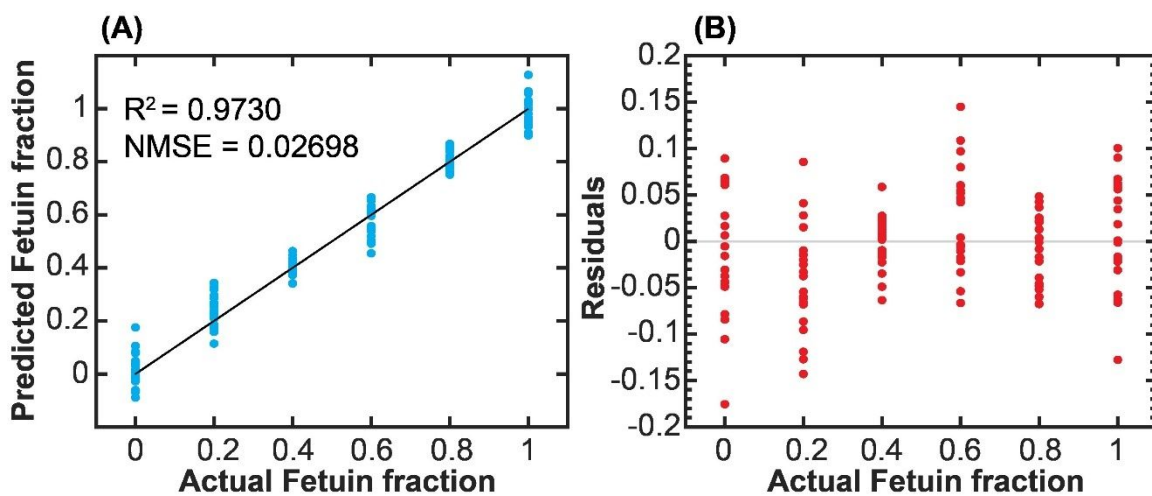
		Collective-spectra				
True Class	Cow	200				
	Goat		200			
	Soy			200		
	Almond				200	
	Oat					200
Specificity		100.0%	100.0%	100.0%	100.0%	100.0%
Sensitivity		100.0%	100.0%	100.0%	100.0%	100.0%
		Cow	Goat	Soy	Almond	Oat
		Predicted Class				

SI Figure 14. Confusion matrix on collective spectra (4MBA, 1TBA, 3MBA, 4ABA, PyriBA) of whole milk classification (100% accuracy).



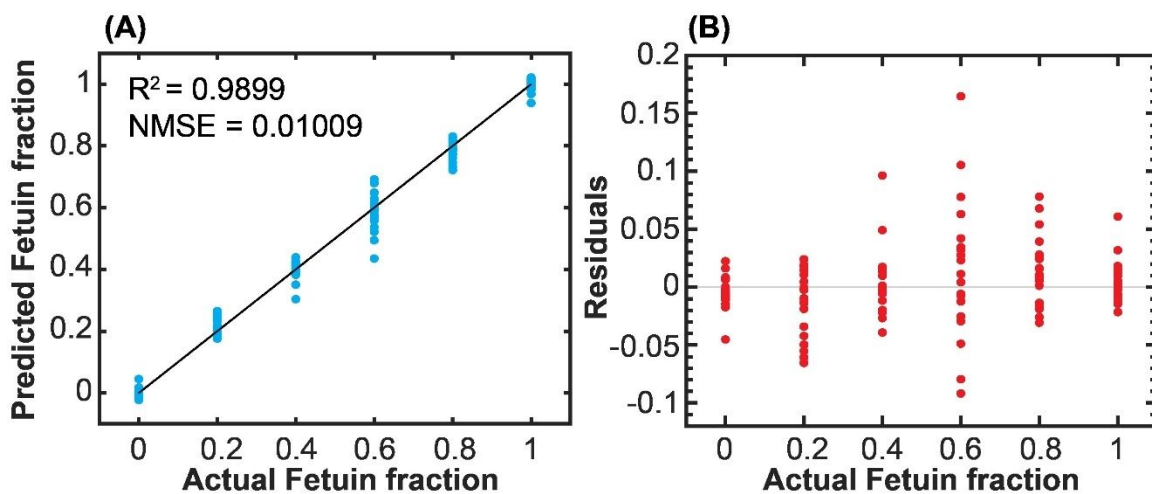
SI Figure 15. Average normalized spectra ($n=200$) of fetuin and asialofetuin on 1TBA, 3MBA, 4ABA and PyriBA functionalized substrates.

4MBA



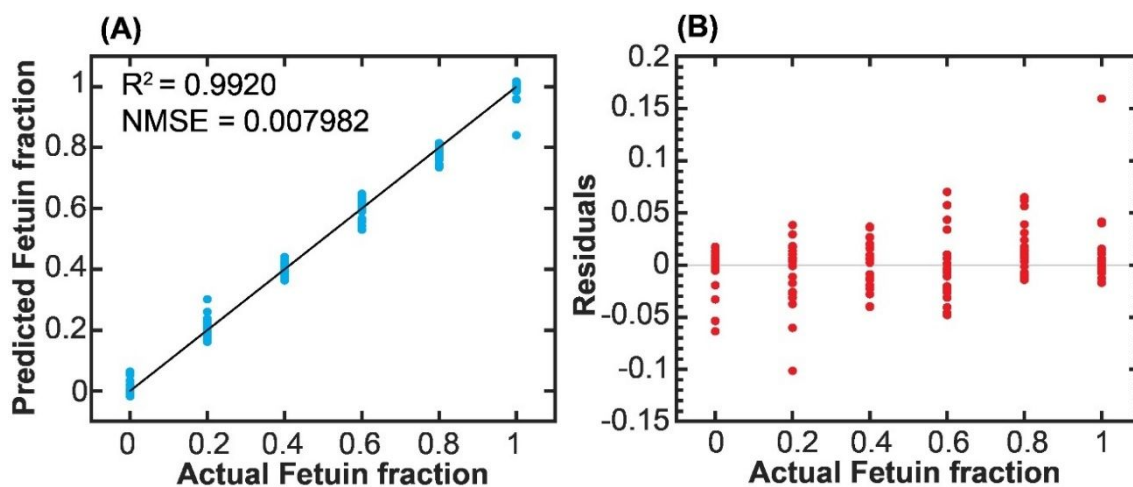
SI Figure 16. Predicted fetuin fraction vs. actual fetuin fraction (A) and residual vs. actual fetuin fraction (B) for 4MBA functionalized substrate.

1TBA



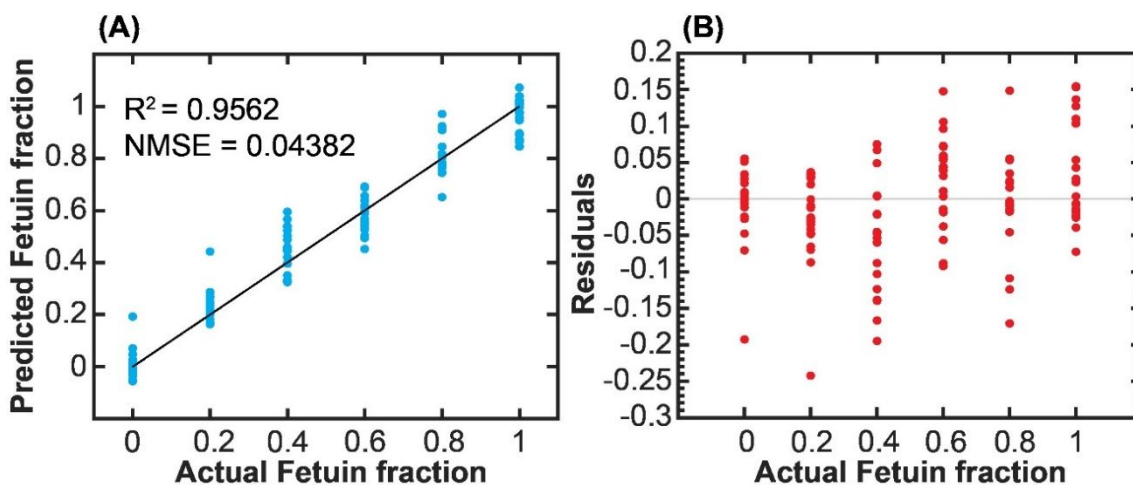
SI Figure 17. Predicted fetuin fraction vs. actual fetuin fraction (A) and residual vs. actual fetuin fraction (B) for 1TBA functionalized substrate.

3MBA



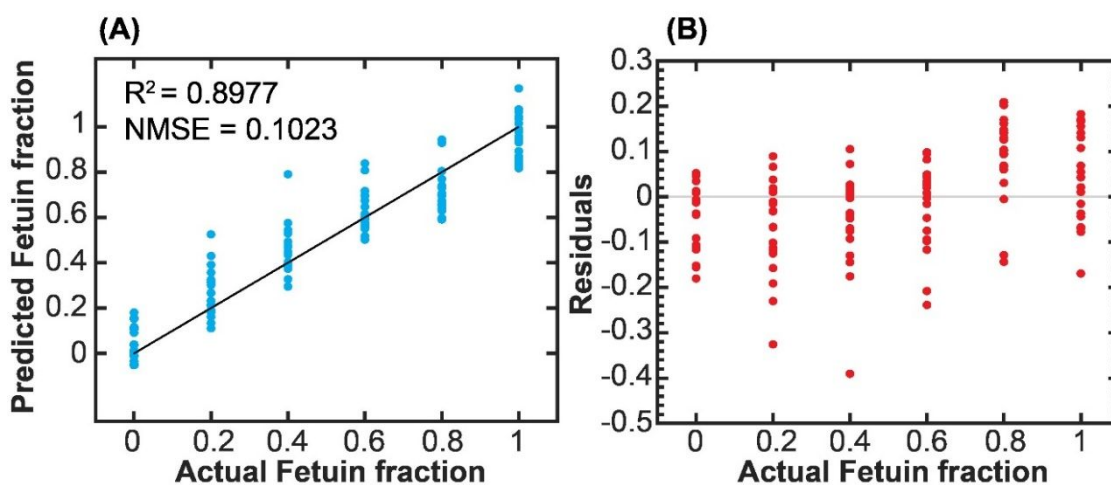
SI Figure 18. Predicted fetuin fraction vs. actual fetuin fraction (A) and residual vs. actual fetuin fraction (B) for 3MBA functionalized substrate.

4ABA



SI Figure 19. Predicted fetuin fraction vs. actual fetuin fraction (A) and residual vs. actual fetuin fraction (B) for 4ABA functionalized substrate.

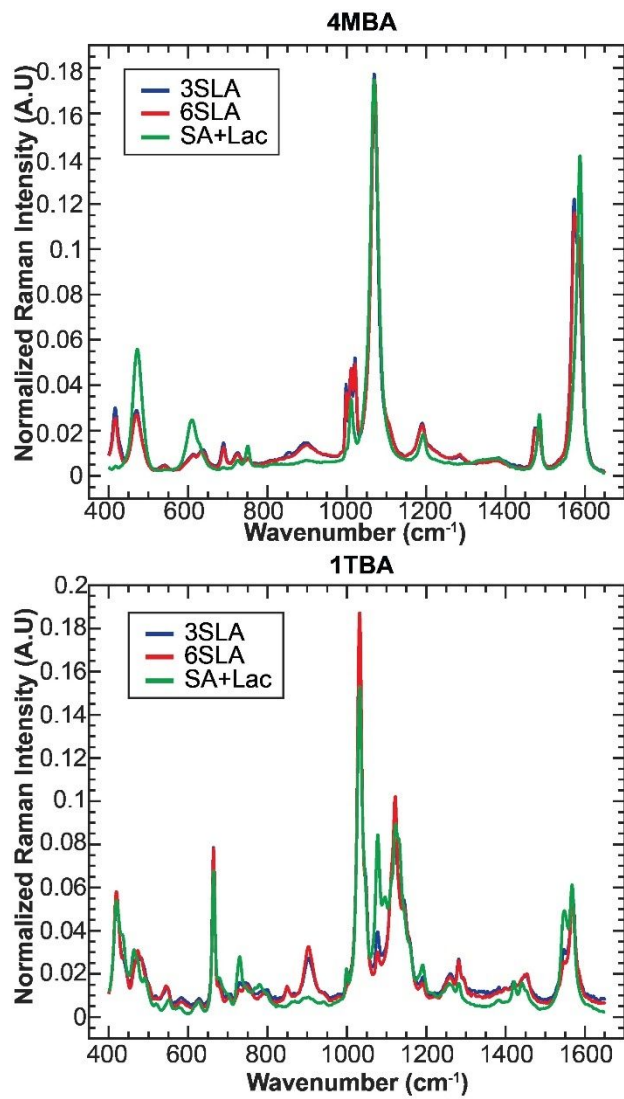
PyriBA



SI Figure 20. Predicted fetuin fraction vs. actual fetuin fraction (A) and residual vs. actual fetuin fraction (B) for PyriBA functionalized substrates.

		4MBA			1TBA		
True Class	SA+Lac	200			200		
	3SLA		197	3		185	15
	6SLA			200		22	178
Sensitivity		100.0%	98.5%	100.0%	100.0%	92.5%	89.0%
Specificity		100.0%	100.0%	99.25%	100.0%	94.5%	96.25%
		SA+Lac	3SLA	6SLA	SA+Lac	3SLA	6SLA
		Predicted Class			Predicted Class		

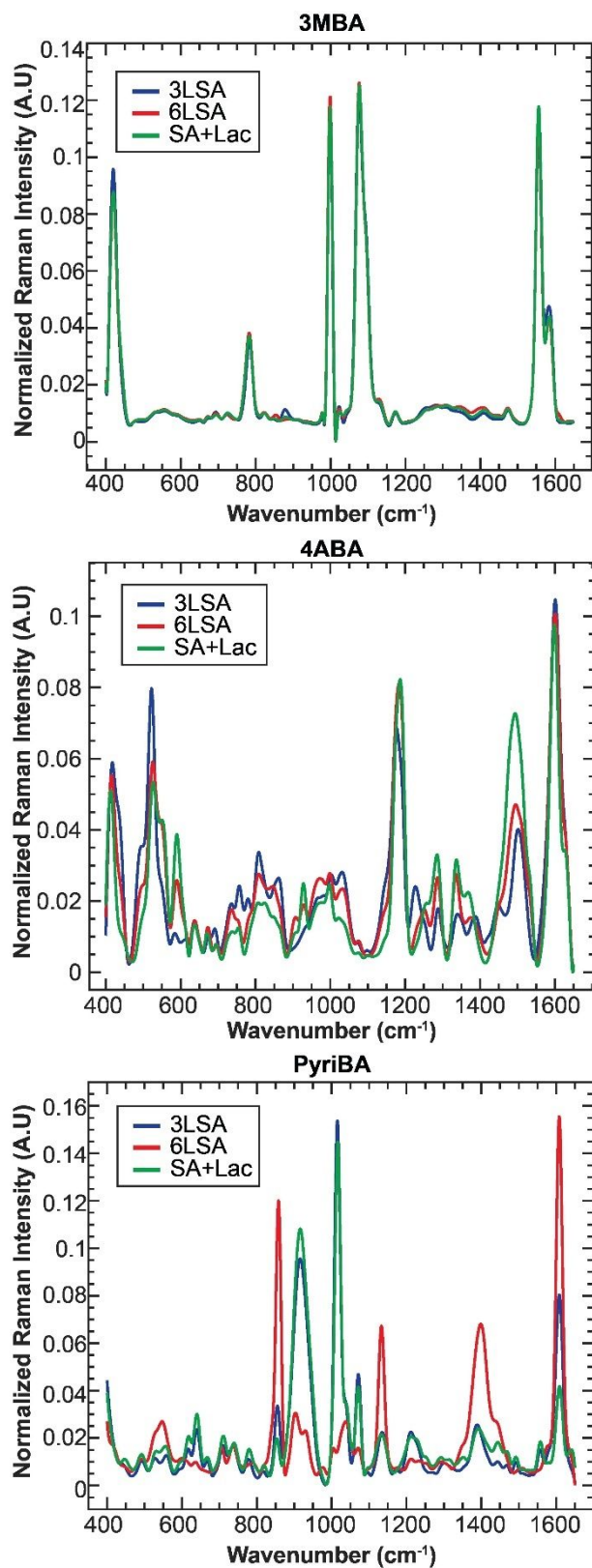
SI Figure 21. Confusion matrix for sialic acid linkage case (SA+Lac, 3SLA and 6SLA) using 4MBA (99.5% accuracy) and 1TBA functionalized substrates (93.8% accuracy).



SI Figure 22. Average normalized spectra ($n = 200$) for 2,3 sialyllactose (3SLA), 2,6 sialyllactose (6SLA), and sialic acid with lactose (SA+Lac) on 4MBA and 1TBA functionalized substrates.

	3MBA			4ABA			PyriBA		
True Class	SA+Lac	200		200		200			
	3SLA		200		200		200		
	6SLA							200	
Sensitivity	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	
Specificity	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	
	SA+Lac	3SLA	6SLA	SA+Lac	3SLA	6SLA	SA+Lac	3SLA	6SLA
	Predicted Class			Predicted Class			Predicted Class		

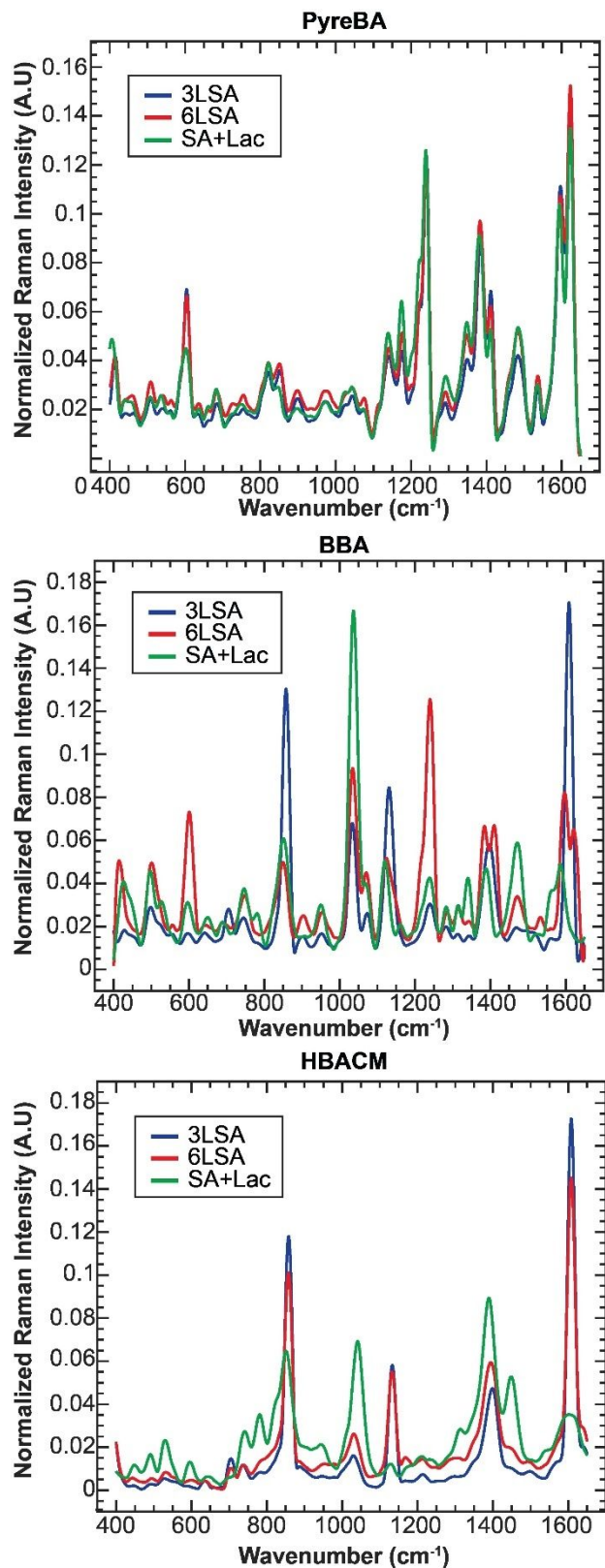
SI Figure 23. Confusion matrix for sialic acid linkage case (SA+Lac, 3SLA and 6SLA) using 3MBA (100% accuracy) 4ABA (100% accuracy) and PyriBA functionalized substrates (100% accuracy).



SI Figure 24. Average normalized spectra ($n = 200$) for 2,3 sialyllactose (3SLA), 2,6 sialyllactose (6SLA), and sialic acid with lactose (SA+Lac) on 3MBA, 4ABA, and PyriBA functionalized substrates.

	PyreBA			HBACM			BBA		
True Class	SA+Lac	199	1	200			200		
	3SLA	1	188	11	199	1	200		
	6SLA	8	192	2	198	200			
Sensitivity	99.5%	94.0%	96.0%	100.0%	99.5%	99.0%	100.0%	100.0%	100.0%
Specificity	99.75%	98.0%	97.0%	100.0%	99.5%	99.75%	100.0%	100.0%	100.0%
	SA+Lac	3SLA	6SLA	SA+Lac	3SLA	6SLA	SA+Lac	3SLA	6SLA
	Predicted Class			Predicted Class			Predicted Class		

SI Figure 25. Confusion matrix for sialic acid linkage case (SA+Lac, 3SLA and 6SLA) using PyreBA (96.5% accuracy) HBACM (99.5% accuracy) and BBA functionalized substrates (100% accuracy).



SI Figure 26. Average normalized spectra ($n = 200$) for 2,3 sialyllactose (3SLA), 2,6 sialyllactose (6SLA), and sialic acid with lactose (SA+Lac) on 3MBA, 4ABA, and PyriBA functionalized substrates.