Supplemental Information

DNAzyme-based biosensor as a rapid and accurate verification tool to complement simultaneous enzyme-based media for *E. coli* detection

Noor Zaouri[1], Zhengfang Cui[2], Ana Sol Peinetti[2], Yi Lu[2], Pei-Ying Hong[1]*

1. King Abdullah University of Science and Technology (KAUST), Water Desalination and Reuse Center, Division of Biological and Environmental Science and Engineering, Thuwal 239455-6900, Saudi Arabia
2. Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801, United States

* Corresponding author

Pei-Ying Hong

Email: peiying.hong@kaust.edu.sa

Tel: +966 12 808 2218

Protein identification using LC-MS/MS

1.1. Protein digestion

The eluted protein solution from the silica based membrane solution was run in SDS-dPAGE to remove impurities that could inhibit protein digestion. Gel bands were excised and cut into small pieces of 1 by 1 mm, and transferred to a sterile 2 mL tube. Gel pieces were de-stained for 5 min with shaking in 100 µL 50% acetonitrile (ACN) in 100 mM ammonium bicarbonate (NH_4HCO_3) for at least 3 times. The gel pieces were dried with 100 µL of ACN for 10 min. Dehydrated samples were reduced by incubation for 20 min at 45°C in 50 µL 10 mM TCEP (tris 2-carboxyethyl phosphine) in 100 mM NH_4HCO_3. After removal of reducing solution, dehydration with ACN was performed, and then 50 µL of alkylating solution (20 mM dimethylacrylamide in 100 mM NH_4HCO_3) was added and incubated for 20 min. Gel pieces were washed twice with the de-staining solution and dried with 100 µL ACN. The samples were digested overnight at 37°C with 50 µL 0.01 µg/µL of trypsin solution in 50 mM NH_4HCO_3 and 5% ACN and pH 8.5. Elution of peptides from the gel pieces was performed by adding 2 µL of formic acid and then 100 µL of ACN. The mixture was then incubated for 10 min with shaking. The supernatant was collected in a new clean tube, and the gel pieces were re-hydrated with 50 µL of 5% formic acid in water, and then 100 µL of ACN was added and the supernatant was collected (two times). All supernatants were combined and dried with a SpeedVac prior purification.

1.2. Protein purification

First, the dried peptides were dissolved with 100 uL of 0.1% trifluoroacetic acid (TFA) in water. Then, OMIX C18 tips) were conditioned twice with 50:50 % (v/v) water and ACN and then twice with 0.1% TFA. The peptides were pipetted in and out of the tip for 10 times. Peptides bound to the tip were cleaned by pipetting with 0.1% TFA twice. The peptides were eluted with 75% ACN then with 0.1% TFA. The eluted solution were collected and dried with SpedVac. The sample was dissolved with 2 % ACN and 0.1% TFA for mass spectrometry (MS /MS) analysis.

1.3. LC-MS/MS operating condition

NanoLC MS/MS analysis was performed using an on-line system consisting of a nano-pump UltiMate 3000 UHPLC binary HPLC system (Dionex, ThermoFisher) coupled to a Q-Exactive HF mass spectrometer (ThermoFisher, Biberach, Germany). Peptides were re-suspended in 3% ACN, 0.1% formic acid and injected into a pre-column 300 μm×5 mm (Acclaim PepMap, 5 μm particle size). After loading, peptides were injected to an Acclaim PepMap100 C18 capillary column (75 μm × 15 cm, 100 Å, 3 μm particle sizes).

Peptides were eluted into the MS at a flow rate of 300 nL/min, using a 40 min gradient from 7% to 38% mobile phase B (80% acetonitrile and 0.1% formic acid). Mobile phase A was 0.1% formic acid in H_2O.

The mass spectrometer was operated in data-dependent mode, with a single MS scan (350-1400 m/z at 60 000 resolution (at 200 m/z) in a profile mode) followed by MS/MS scans on the 10 most intense ions at 15 000 resolution. Ions selected for MS/MS scan were fragmented using higher energy collision dissociation (HCD) at normalized collision energy of 28% with an isolation window of 1.8 m/z.

1.4. Data Analysis

Raw files were converted to MGF files using Protein Discoverer version 2.2 and searched against the *E. coli* UniProt database using MASCOT Version 2.6 (Matrix Science Ltd, London, UK). Search parameters were peptide mass tolerance of 5 ppm (parts per million), and MS/MS tolerance of 0.02 amu allowing 2 missed cleavage. Dimethyl-Propionamide (C) was set as fixed modification, and acetyl (Protein N-term), and Oxidation (M), were allowed as variable modification. Peptide assignments with ion score cut-off of 20 and a significance threshold of $\rho < 0.05$ were exported to Excel for further analysis.

To narrow down the protein list that hybridized to the liquid sensor, only ubiquitous proteins that were consistently present in all tested CIM were considered. In addition, protein targets must be bigger than 30 kDa.

Table S1. List of the investigated fecal coliforms and total coliforms bacteria (target and non-target) reported by Zhang, et al. [1], and tested against DNAzyme biosensor in this study.

Bacteria name	ID	Bacteria name	ID	Bacteria name	ID	Bacteria name	ID
E. coli	OH 1	*Enterobacter* VIII	FL-10-SCAN -4	*Serratia group* II	2-EPA 2	*Aeromonas*	021110-12
	OH 3	*Wautersia*	3530		EPA 75	*Acinetobacter*	PA-15-CULT-3
	OH 11	*Enterobacter* IX	EPA 72		EPA 76		PA-15-CULT-9
	OH 18		012610-08		EPA 214		PA-15-CULT-10
***Klebsiella* I**	OH 2		612610-12	*Providencia*	OH 10		PA-16-CULT-5
	OH 8		012610-06		CCP-1-3	*Chryseobacterium*	VA-5-CULT-4
	OH 15		02044-07		6-1		VA-6-CULT-4
	EPA 77		02664-06		10-2		FL-7-CULT-4
	EPA 185	*Citroabacter* II	OH 6		006		NJ-20-CULT-6
	EPA 193		EPA 81		008	*Micobacterium*	FL-9-CULT-2
***Enterobacter* V**	EPA 59		EPA 153	*Morganell*	CA-33-Colite-3		FL-9-CULT-3
	EPA 61		EPA 159		MA-35-Ready 2		FL-9 CUTL-4
	EPA 176		EPA 191		VA-53-Ready-2		FL-9-CULT-5
	EPA 179		EPA 192		VA-54-colite-2		
	EPA 180	*Raoultella* group I	15		EPA 227		
	EPA 181		121		007		
***Enterobacter* VI**	EPA 64		EPA 10	*Plesiomonas*	EPA 45		
	EPA 69		EPA 117		EPA 49		
	EPA82		EPA 194		EPA 50		
	EPA86	*Serratia* group I	EPA 74		EPA 52		
	EPA186		EPA 231		EPA 55		
	EPA187		EPA 244		EPA 56		
***Enterobacter* VIII**	FL-10-BLUE-3		EPA 246	*Aeromonas*	EPA 124		
	FL-10-BLUE-4		EPA 248		EPA 131		
	FL-10-BLUE 5		20410-17		EPA 150		
	FL-10-CULT-2	*Serratia* group II	CO-1-CULT-1		011110-09		
	FL-10-SCAN-2		MA-21-READY		021110-11		

Table S2. List of proteins in *E. coli* OH11 CIM that interacted with *E. coli* DNAzyme.

Protein	Description	Protein score	MW (KDa)	Protein match	Protein sequence
P04949	Flagellin	3949	51.3	93	DDAAGQAIANR LSSGLRINSAKDDAAGQAIANR INSAKDDAAGQAIANR AQVINTNSLSLITQNNINK
P0A8T7	DNA-directed RNA polymerase subunit beta	685	155.1	27	DLLGITK LIDEFGR VADLFEAR GLKENVIVGR SVITVGPYLR LLDLAAPDIIVR FATSDLNDLYRR QLNVFEGERVER LIPAGTGYAYHQDR IGLLLDMPLRDIER EGLNVLQYFISTHGAR YNKVIDIWAAANDRVSK GLMAKPDGSIIETPITANFR
P46889	DNA translocase	40	146.6	3	SLSTVAVR
P06612	DNA topoisomerase	20	97.3	1	ALVIVESPAK
P0AES4	DNA gyrase subunit A	51	96.9	1	NTQGVILIR
P36659	Curved DNA-binding protein	26	34.4	1	QTGDLYAVLK
P46850	RNA-splicing ligase	17	45.2	7	KLFSVEDQIR
P75937	Flagellar hook protein	923	42.0	21	ELVNMIVAQR

P0A6Y8	Chaperone protein	1396	69.1	42	DVSIMPFK VAEFFGKEPR MAPPQISAEVLK KVAEFFGKEPR MAPPQISAEVLKK AKLESLVEDLVNR SLGQFNLDGINPAPR QAVTNPQNTLFAIKR IIAADNGDAWVEVKGQK AADNKSLGQFNLDGINPAPR IINEPTAAALAYGLDKGTGNR HSQVFSTAEDNQSAVTIHVLQGER AKIELSSAQQTDVNLPYITADATGPK TAEDYLGEPVTEAVITVPAYFNDAQR KTAEDYLGEPVTEAVITVPAYFNDAQR VLENAEGDRTTPSIIAYTQDGETLVGQPAKR
P0A7Z4	DNA-directed RNA polymerase subunit alpha	851	36.5	27	GLSLGMR TEVELLK VTLEPLER IAYNVEAAR GFGHTLGNALR SLTEIKDVLASR MQGSVTEFLKPR LVDIEQVSSTHAK VQGKDEVILTLNK KSLTEIKDVLASR LLVDACYSPVER AEAIHYIGDLVQR IAYNVEAARVEQR
					LVDSNGSVFYSR SGTASFADMFAGSK GLDVAISQNGFFR TQDQILNTLVNLR VAGITQDFTDGTTTNTGR GSVTVFDSQGNAHDMSVYFVK AFSQAVSGLNAAATNLDVIGNNIANSATYGFK

P33235	Flagellar hook-associated protein	580	57.9	16	SEGLVNQFK QIASLNDQISR QLAAVPSSADPSR EYDAFITNQLR LLNTGSLGGILTFR ISFDNNQWQVTR LASNTTFTVTPDANGK SFNDAYASLVSDIGNK TTVAYVDGTAGNIEIPEK AGFDANGDAGEDFFAIGKPAVLQNTK TTDQYLRDQDKQVNIAIGASVDQINNYAK

Table S3. List of proteins in *E. coli* OH18 CIM that interacted with *E. coli* DNAzyme.

Protein ID	Description	Protein score	MW (KDa)	protein match	Protein sequence
P04949	Flagellin	2491	51.3	32	LSSGLRINSAK FTSNIKGLTQAAR AQIIQQAGNSVLAK INSAKDDAAGQAIANR IQDADYATEVSNMSK AQVINTNSLSLITQNNINK LSSGLRINSAKDDAAGQAIANR INSAKDDAAGQAIANRFTSNIK IQDADYATEVSNMSKAQIIQQAGNSVLAK
P0A8T7	DNA-directed RNA polymerase subunit beta	447	155.1	15	DLLGITK VADLFEAR GLKENVIVGR SVITVGPYLR LLDLAAPDIIVR FTDMIDGQTITR LGIQAFEPVLIEGK LIPAGTGYAYHQDR EGLNVLQYFISTHGAR
P46889	DNA translocase	47	146.6	3	SLSTVAVR
P06612	DNA topoisomerase	103	97.3	3	LYQLIWR ALVIVESPAK GIGRPSTYASIISTIQDR
P0AES4	DNA gyrase subunit A	209	96.9	5	NTQGVILIR TAEDENVVGLQR GRPIVNLLPLEQDER AGDDAARPEWLEPEFGVR

P0A6Y8	Chaperone protein	1168	69.1	34	IAGLEVK VAEFFGKEPR DAGRIAGLEVKR ASSGLNEDEIQK KVAEFFGKEPR MAPPQISAEVLKK AKLESLVEDLVNR SLGQFNLDGINPAPR QAVTNPQNTLFAIKR AADNKSLGQFNLDGINPAPR IINEPTAAALAYGLDKGTGNR RIINEPTAAALAYGLDKGTGNR HSQVFSTAEDNQSAVTIHVLQGER TAEDYLGEPVTEAVITVPAYFNDAQR KTAEDYLGEPVTEAVITVPAYFNDAQR VLENAEGDRTTPSIIAYTQDGETLVGQPAKR
P0A7Z4	DNA-directed RNA polymerase subunit alpha	615	36.5	20	IAYNVEAAR SLTEIKDVLASR MQGSVTEFLKPR LVDIEQVSSTHAK IHSEEDERPIGR KSLTEIKDVLASR LLVDACYSPVER IAYNVEAARVEQR
P36659	Curved DNA-binding protein	129	34.4	4	QTGDLYAVLK KYHPDVSKEPDAEAR FKEVAEAWEVLSDEQR
P46850	RNA-splicing ligase	17	45.2	7	KLFSVEDQIR
P75937	Flagellar hook protein	710	42.0	12	ELVNMIVAQR TQDQILNTLVNLR VAGITQDFTDGTTTNTGR NYQSNAQTIKTQDQILNTLVNLR AFSQAVSGLNAAATNLDVIGNNIANSATYGFK

P33235	Flagellar hook-associated protein	393	57.9	11	SEGLVNQFK QIASLNDQISR EYDAFITNQLR LLNTGSLGGILTFR ISFDNNQWQVTR LTGVGAGASPNNLLDQR SSLINNAMSGLNAAQAALNTASNNISSYNVAGYT R