Enzymatic fingerprinting of structurally similar homologous proteins using polyion complex library constructed by tuning PEGylated polyamine functionalities

Shunsuke Tomita,[*a] Tomohiro Soejima[b], Kentaro Shiraki[c] and Keitaro Yoshimoto[*a]

[a] Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro, Tokyo, 153-8902, Japan

E-mail: s_tomita@bio.c.u-tokyo.ac.jp; ckeitaro@mail.ecc.u-tokyo.ac.jp

[b] College of Arts and Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro, Tokyo, 153-8902, Japan

[c] Faculty of Pure and Applied Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan

Experimental

Materials

Polymer synthesis: Commercial tetrahydrofuran (THF) (Kanto Chemicals, Tokyo, Japan), 2-(*N*,*N*-dimethylamino)ethyl methacrylate (AMA) (Wako Pure Chemical Ind., Osaka, Japan), and poly(ethylene glycol) methyl ether with number-averaged molecular weight (M_n) of 5000 (Sigma Chemical Co., St. Louis, MO) were purified using conventional methods. Iodoethane (99%), 2-iodoethanol (99%), and benzyl bromide (98%) were also purchased from Sigma Chemical Co., and used without further purification. Potassium naphthalene was prepared as described previously.[1] **Protein analysis:** β-Galactosidase from *Aspergillus oryzae* (GAO), *o*-Nitrophenyl-β-D-galactopyranoside (oNPG), immunoglobulin G from human serum (IMM), α_1-antitrypsin from human plasma (ANT), fibrinogen from human plasma (FIB), apotransferrin from human (TRA), albumin from human serum (HALB), albumin from bovine serum (BALB), albumin from rabbit serum (RALB), and 3-(*N*-Morpholino)propanesulfonic acid (MOPS) were obtained from Sigma Chemical Co. Albumin from equine serum (EALB) was obtained from Rockland Immunochemicals Inc. (Gilbertsville, PA). Poly(ethylene glycol)-*block*-poly(*N,N*-dimethylaminoethyl methacrylate) (PEG-*b*-PAMA, **P1**) with M_n (PEG) 5000 and M_n (PAMA) 12000 was obtained from Polymer Source inc. (Dorval, QC, Canada). All chemicals used were of high-quality analytical grade and were used as received.

*Synthesis of PEG-*b*-PAMA (**P2**)*

PEG-*b*-PAMA with smaller molecular weight (**P2**) was synthesized according to the previously described method with slight modifications.[2] Poly(ethylene glycol) methyl ether (1.00 mmol) and potassium naphthalene (1.00 mmol) were mixed with 40 mL of THF in a 100-mL flask with a three-way stopcock under a nitrogen atmosphere, followed by addition of AMA (30.0 mmol) and stirred for 1 h. The reaction mixture was poured into cold 2-propanol. The obtained precipitate was dissolved in methanol, dried in vacuo, and freeze dried after protonation of the PAMA segment in the block copolymer. Finally, the unreacted poly(ethylene glycol) methyl ether was removed from the sample by Soxhlet extraction with THF (Figure S1). The molecular weights of the PEG and PAMA segments were 4500 (size exclusion chromatography) and 5500 (^{1}H NMR) (Figure S2), respectively.

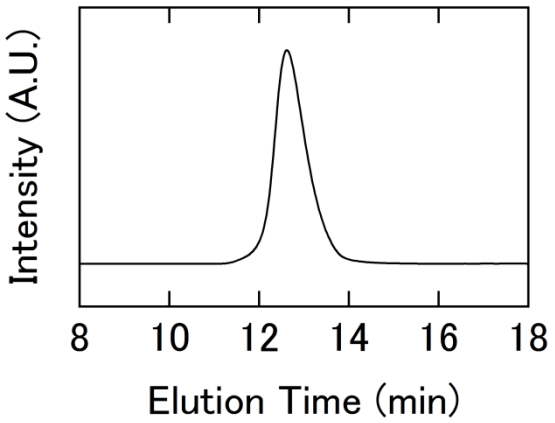

Figure S1. Size exclusion chromatogram of **P2** after Soxhlet extraction (M_w/M_n = 1.31).

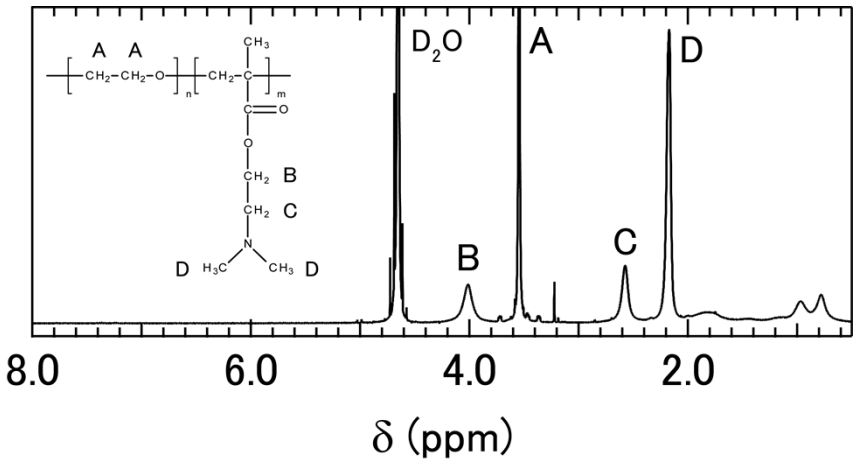

Figure S2. ^{1}H NMR spectrum of **P2** (400 MHz, D_2O).

Synthesis of quaternized PEG-b-*PAMA (PEG*-b-*QPAMA) (**P3**-**P5**)*

Quaternization reaction of the tertiary amino groups of PAMA segment was carried out using functional halides based on previous reports.[3] Briefly, PEG-*b*-PAMA (**P2**) (500 mg, 1.75 mmol amino groups) was dissolved in methanol (10 mL), followed by the addition of alkyl halides (35.0 mmol, 20 molar eq. vs. the amino groups in **P2**) (iodoethane (**P3**), 2.8 mL; 2-iodoethanol (**P4**), 2.7 mL; benzyl bromide (**P5**), 4.2 mL). The reaction mixture was stirred for 6 days at 50°C shielded from the light, dried in vacuo, and poured into cold 2-propanol. The obtained precipitate was solubilized in methanol and precipitated repeatedly with cold 2-propanol. Quantitative quaternization was confirmed by potentiometric titration (Figure S3), according to the previously described procedure.[4] Titration was carried out with 10 mM NaOH at 25°C in the presence of 10 mM of NaCl. **P2** possessed a buffering capacity at pH around 7, while no buffering region was observed for **P3**, **P4**, or **P5** because of the quantitative quaternization of amino groups in **P2** (Figure S3). 99% (**P3**), 97% (**P4**), and 100% (**P5**) of quaternization of the amino groups of **P2** were further determined by ^{1}H NMR (Figure S4).

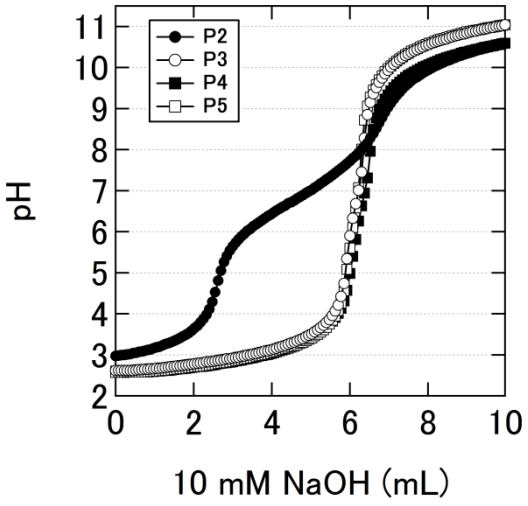

Figure S3. Potentiometric titration curves of the **P2** (closed circles), **P3** (open circles), **P4** (closed squares), and **P5** (open squares) at 25°C in the presence of 10 mM NaCl.

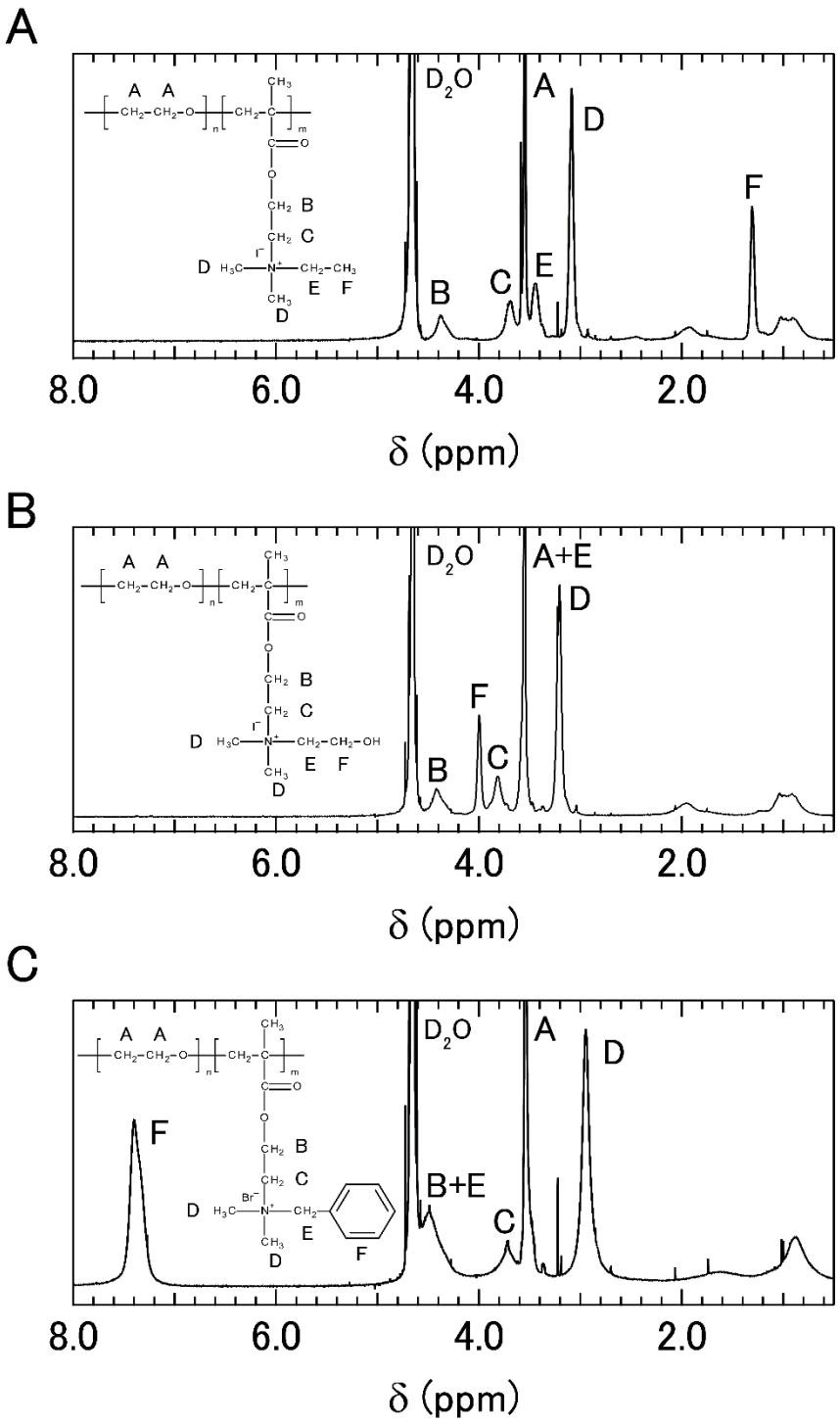

Figure S4. ^{1}H NMR spectra of (A) **P3**, (B) **P4**, and (C) **P5** (400 MHz, D_2O).

*Determination of the isoelectric point (p*I*) of proteins*

The isoelectric point (p*I*) of proteins can be predicted theoretically by considering the p*K* values of the side chains of amino acids (see Figure S7).[5] However, differences are usually observed between predicted and experimentally determined p*I*, as the prediction is based on

the assumption that all ionizable amino acids are accessible to water. Therefore, in this study, the p*I* values of proteins were determined from the dependence of the zeta potential of proteins on pH to precisely evaluate the effects of surface charges of proteins on pattern generation. The zeta-potential measurements of proteins were performed using the MPT-2 Autotitrator (Malvern Instruments, Ltd., Malvern, Worcestershire, UK) in parallel with a Zetasizer Nano ZS (Malvern Instruments, Ltd.). Basically, 1.0 mg/mL protein in 10 mM MOPS was titrated from acidic pH (pH = 4.0) to basic pH (pH = 8.5) utilizing 0.1 M HCl and 0.1 M NaOH at 25°C. At every 0.5 pH unit (± 0.05) the zeta potential was determined, and p*I* was calculated using Zetasizer Software Version 7.02 (Figures S5 and S7).

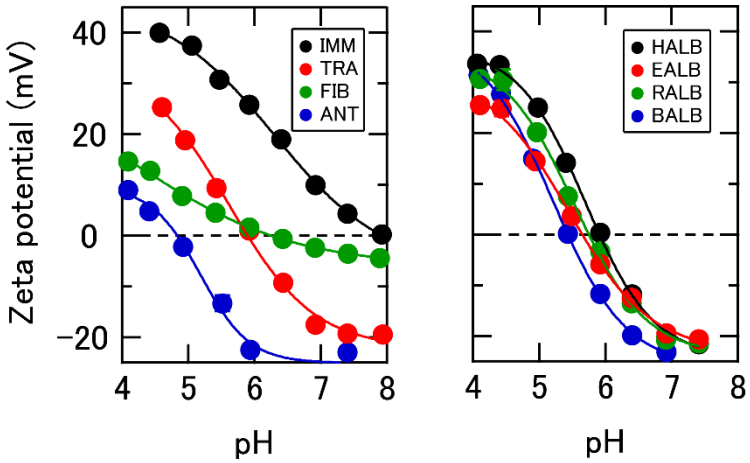

Figure S5. Zeta potential of proteins as a function of pH at 25°C in the presence of 10 mM MOPS.

Determination of GAO and protein concentrations

Concentrations of GAO and proteins were determined from the absorbance at 280 nm using a spectrophotometer (UV-2450; Shimadzu Corporation, Kyoto, Japan), with extinction coefficients of 192075 M^{-1} cm^{-1} (GAO), 1.427 $(mg/mL)^{-1}$ cm^{-1} (IMM), 0.450 $(mg/mL)^{-1}$ cm^{-1} (ANT), 1.689 $(mg/mL)^{-1}$ cm^{-1} (FIB), 1.132 $(mg/mL)^{-1}$ cm^{-1} (TRA), 0.518 $(mg/mL)^{-1}$ cm^{-1} (HALB), 0.646 $(mg/mL)^{-1}$ cm^{-1} (BALB), 0.657 $(mg/mL)^{-1}$ cm^{-1} (RALB), and 0.501 $(mg/mL)^{-1}$ cm^{-1} (EALB).[6]

Titration of PEGylated polyamines to GAO

A solution containing 12 nM GAO in 10 mM MOPS (pH 7.0) was prepared and cryopreserved at –80°C. Frozen GAO solution was thawed immediately before experiments in a 37°C water bath. Only one freeze-thaw cycle did not significantly affect enzyme activity (Unfrozen, 8.6 ± 0.9 nM s^{-1}; Frozen, 8.1 ± 0.4 nM s^{-1}). Various concentrations of PEGylated

polyamines were incubated with 1.25 nM GAO in 10 mM MOPS (pH 7.0). After incubation for 30 min at 25°C, 120 μL of each solution was loaded into each well of 96-well plates (Half Area 96-Well Clear Flat Bottom UV-Transparent Microplates; Corning Inc., Corning, NY). Subsequently, 30 μL of 25 mM oNPG in 10 mM MOPS (pH 7.0) was added to each well so that the final concentrations were 1.0 nM GAO and 5 mM oNPG. The time course of the increase in absorbance at 400 nm was then recorded using a microplate reader (Viento® nano; DS Pharma Biomedical Co. Ltd., Osaka, Japan). The samples were measured in triplicate.

*Titration of proteins to GAO/**P3** complex*

Aliquots of 100 μL of solution containing 1.5 nM GAO, 37.5 nM **P3**, and 10 mM MOPS (pH 7.0) were loaded into each well of 96-well plates. Subsequently, 20 μL of various concentrations of IMM, FIB, and HALB was added to each well. After incubation for 30 min at 25°C, 30 μL of 25 mM oNPG in 10 mM MOPS (pH 7.0) was further added to each well so that the final concentrations were 1.0 nM GAO, 25 nM **P3**, and 5 mM oNPG. The time course of the increase in absorbance at 400 nm was then recorded using a microplate reader (Figure S6). The samples were measured in triplicate.

Sensing of protein analytes

GAO and PEGylated polyamines were mixed at the optimal ratio determined taking into account both differences in activity changes between proteins (1.5 nM GAO with 37.5 nM **P1**, **P2**, **P3**, and **P4** or 30 nM **P5**). Aliquots of 100 μL of each solution were loaded into each well of 96-well plates. Subsequently, 20 μL of protein analytes (Figure S7) was added. After incubation for 30 min at 25°C, 30 μL of 25 mM oNPG in 10 mM MOPS (pH 7.0) was further added to each well so that the final concentrations were 1.0 nM GAO with 25 nM **P1**, **P2**, **P3**, and **P4** or 20 nM **P5**. The time course of the increase in absorbance at 400 nm was then recorded using a microplate reader for 20 min. This process was repeated for protein analytes with 5 EPCs in six replicates each. This data set matrix was subjected to linear discriminant analysis (LDA) using SYSTAT 13 (Systat Inc., Evanston, IL). Similar procedures were also performed for discrimination of unknown protein analytes.

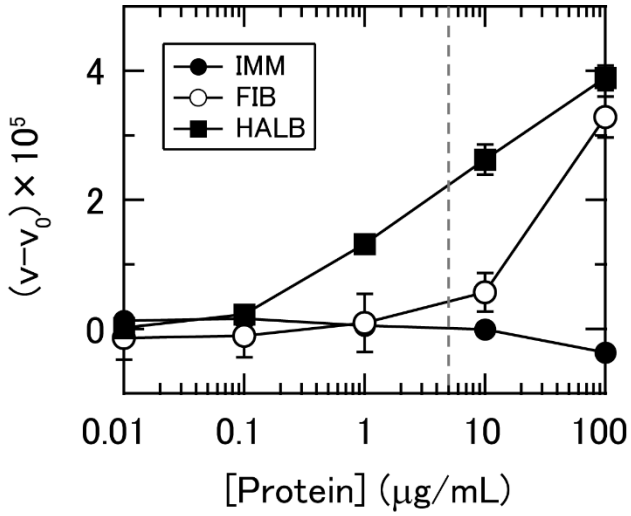

Figure S6 Changes in activity of PIC of GAO/**P3**. Titration of proteins to 1.0 nM GAO in 10 mM MOPS (pH 7.0). The *y*-axis indicates that changes in initial slope of Abs_{400} derived from enzymatic hydrolysis of substrates (v-v_0).

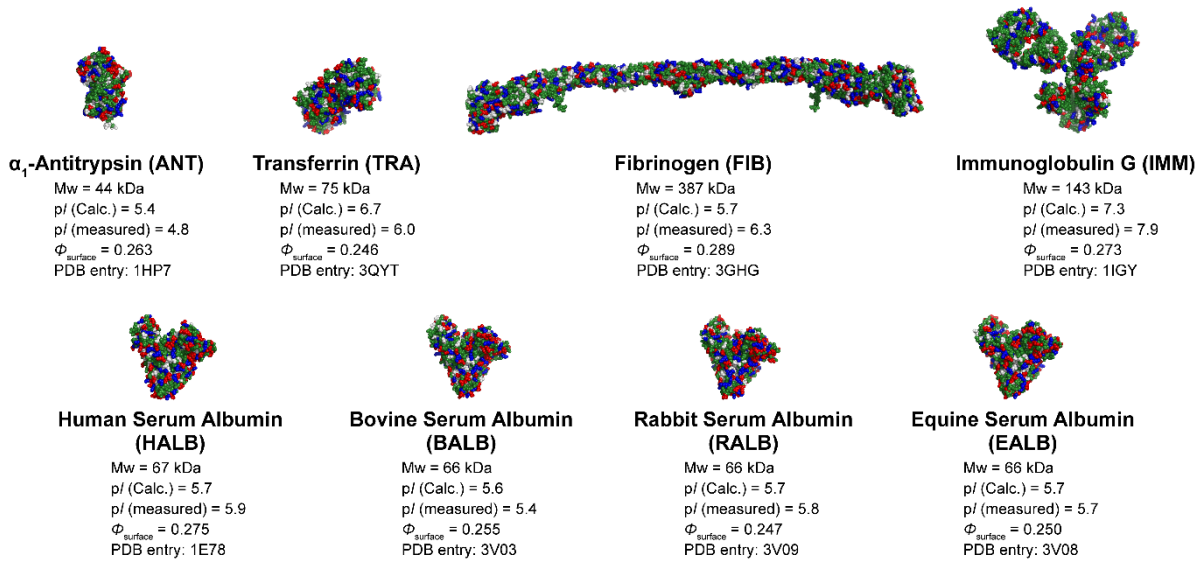

Figure S7 Properties of proteins used in this study. Color scheme for proteins: anionic residues, red; cationic residues, blue; hydrophobic residues, gray; hydrophilic residues, green. p*I* (Calc.) values were obtained theoretically taking into account the amino acid composition of proteins and the p*K* values of the side chains. p*I* (measured) values were obtained from the pH dependence of zeta-potential (see Experimental section). The surface hydrophobicity ($\Phi_{surface}$) of proteins based on the Miyazawa–Jernigan hydrophobicity scale was estimated using the previously described method[7] with the accessible surface areas of proteins calculated by the program GETAREA.[8]

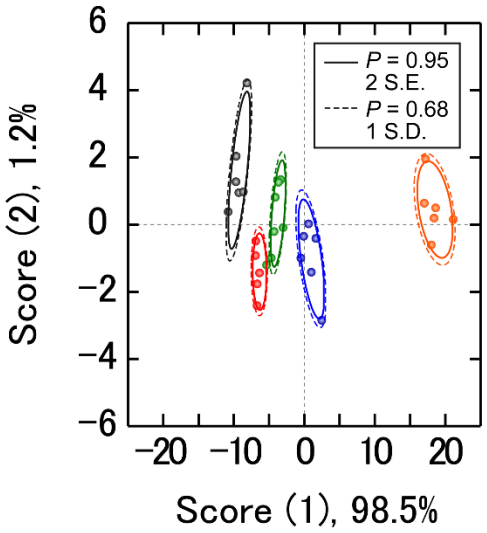

Figure S8 Discriminant score plot of the first two discriminant functions of enzyme activity patterns for five human plasma proteins using three PICs (same as Fig. 3) analyzed by LDA. The ellipses representing confidence intervals ($P = 0.95$, ± 2 standard error; solid lines) are slightly smaller than confidence intervals ($P = 0.68$, ± 1 standard deviation; broken lines) when $n = 6$.

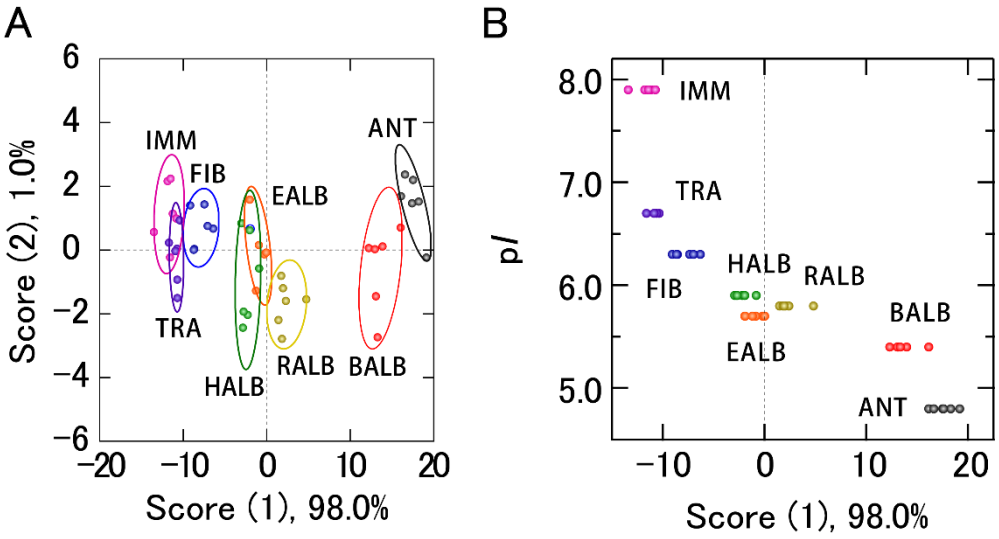

Figure S9 Pattern-based sensing of proteins using five PICs (GAO with **P1**, **P2**, **P3**, **P4**, and **P5**). (A) Discriminant score plot of the first two discriminant functions of enzyme activity patterns analyzed by LDA. Raw data of enzyme activity patterns are shown in Table S1. The ellipses represent confidence intervals ($P = 0.68$, ± 1 standard deviation) for the individual proteins. Accuracy of 92% was obtained *via* the Jackknife classification. (B) First discriminant scores vs. p*I* values of proteins.

Table S1. Data set matrix of $(v\text{-}v_0)\times10^5$ for individual protein analytes generated from the sensor array containing 5 PICs

Analytes	GAO/**P1**	GAO/**P2**	GAO/**P3**	GAO/**P4**	GAO/**P5**
ANT	2.871	2.545	3.482	3.590	3.056
ANT	2.894	2.540	3.194	3.331	2.912

ANT	3.199	2.593	3.493	3.382	3.226
ANT	2.827	2.964	3.570	3.776	3.432
ANT	2.867	2.457	3.570	3.753	2.951
ANT	2.731	2.443	3.466	3.585	2.603
FIB	–0.036	0.111	0.409	1.239	0.750
FIB	0.435	0.442	0.690	0.752	0.737
FIB	0.156	0.294	0.649	1.115	0.572
FIB	0.019	0.248	0.434	0.813	0.808
FIB	–0.253	–0.085	0.744	0.980	0.737
FIB	–0.148	0.129	0.981	1.203	0.473
IMM	–0.146	0.059	0.060	0.003	0.268
IMM	0.012	–0.247	0.145	0.244	–0.003
IMM	–0.331	–0.315	–0.054	–0.014	0.086
IMM	0.018	–0.035	0.122	0.240	0.230
IMM	0.303	0.034	–0.062	0.042	0.026
IMM	–0.096	0.342	0.077	0.275	–0.146
TRA	0.091	–0.098	0.233	0.588	0.751
TRA	–0.218	0.052	0.216	0.361	0.721
TRA	0.082	0.012	0.087	0.378	0.870
TRA	–0.403	0.126	0.207	0.609	0.884
TRA	–0.198	–0.258	0.142	0.579	0.827
TRA	–0.574	–0.313	0.562	0.764	0.411
HALB	0.451	0.912	1.245	1.073	1.414
HALB	0.009	0.650	1.424	1.471	1.502
HALB	0.632	0.437	1.301	1.393	1.229
HALB	0.564	0.542	1.670	1.772	1.772
HALB	0.876	0.864	1.125	1.494	1.464
HALB	0.329	0.941	1.154	1.230	1.388
BALB	2.409	2.525	2.962	2.850	2.547
BALB	1.928	2.296	3.030	3.195	3.027
BALB	2.281	2.128	3.026	2.011	1.881
BALB	2.685	2.580	3.250	3.294	2.915
BALB	2.333	2.372	2.811	3.055	2.908
BALB	1.876	2.734	2.843	2.951	2.937
EALB	1.014	0.867	1.091	1.495	1.288
EALB	0.898	1.228	1.295	1.570	1.444
EALB	0.969	1.321	1.284	1.443	1.314
EALB	0.949	1.226	0.922	1.342	1.185
EALB	1.008	0.946	1.277	1.353	1.528
EALB	0.877	0.867	1.323	0.930	1.545
RALB	1.083	1.396	1.616	1.451	1.665
RALB	0.679	1.316	1.642	1.825	1.943
RALB	0.791	1.465	1.674	1.740	1.304
RALB	1.099	1.505	2.122	1.998	1.982
RALB	1.034	1.727	1.452	1.790	1.889
RALB	0.998	1.882	1.332	1.167	1.623

Table S2. Classification accuracy of sensor arrays for discrimination of 5 human plasma proteins

	Selected PICs					%correct					
	GAO/**P1**	GAO/**P2**	GAO/**P3**	GAO/**P4**	GAO/**P5**	HALB	ANT	FIB	IMM	TRA	Total
1 EPC						83	100	33	17	67	60
						83	100	67	50	50	70
						100	100	50	83	50	77
						83	100	50	100	83	83
						100	100	33	100	83	83
2 EPCs						83	100	67	17	50	63
						100	100	83	100	50	87
						67	100	100	100	83	90
						100	100	83	100	67	90
						100	100	83	67	67	83
						100	100	100	83	100	97
						100	100	83	100	67	90
						100	100	67	100	67	87
						100	100	67	100	83	90
						100	100	67	100	83	90
3 EPCs						100	100	83	50	50	77
						100	100	100	83	83	93
						100	100	83	100	100	97
						100	100	100	100	67	93
						100	100	67	100	83	90
						100	100	100	100	100	**100**
						100	100	100	83	83	93
						100	100	83	100	83	93
						100	100	100	100	100	**100**
						100	100	83	100	83	93
4 EPCs						100	100	100	83	83	93
						100	100	83	100	83	93
						100	100	83	100	100	97
						100	100	100	100	83	97
						100	100	100	100	83	97
5 EPCs						100	100	100	100	100	**100**

Table S3. Discrimination of 20 unknown human plasma proteins by the PIC sensor array consisting of GAO/**P2**, GAO/**P4** and GAO/**P5**

Identification	Enzyme activity pattern			Verification	Accuracy
	GAO/**P2**	GAO/**P4**	GAO/**P5**		
HALB	1.223	1.463	1.567	HALB	YES
HALB	0.969	1.234	1.495	HALB	YES
HALB	0.935	1.446	1.544	HALB	YES
HALB	0.909	1.700	1.477	HALB	YES
IMM	0.301	0.254	–0.050	IMM	YES
IMM	0.303	0.034	0.158	IMM	YES
IMM	0.090	–0.051	0.050	IMM	YES
IMM	0.216	0.097	–0.219	IMM	YES
TRA	0.081	0.447	0.551	TRA	YES
TRA	0.219	0.443	0.659	TRA	YES
TRA	–0.064	0.486	0.530	TRA	YES
TRA	0.048	0.466	0.622	TRA	YES
FIB	0.499	0.962	0.723	FIB	YES
FIB	0.308	0.964	0.697	FIB	YES
FIB	0.209	1.238	0.571	FIB	YES
FIB	0.614	0.810	0.738	FIB	YES
ANT	2.717	3.258	3.339	ANT	YES
ANT	2.777	3.478	3.277	ANT	YES
ANT	2.535	3.582	3.129	ANT	YES
ANT	2.442	3.054	3.116	ANT	YES

Table S4. Classification accuracy of sensor arrays for discrimination of 4 homologous albumins

	Selected PICs					%correct				
	GAO/**P1**	GAO/**P2**	GAO/**P3**	GAO/**P4**	GAO/**P5**	BALB	EALB	HALB	RALB	Total
1 EPC						100	17	83	0	50
						100	50	67	83	75
						100	33	33	67	58
						83	33	50	67	58
						83	50	33	83	63
2 EPCs						100	83	83	83	88
						100	100	67	83	88
						100	83	83	50	79
						100	83	83	83	88
						100	67	67	100	83
						83	50	67	100	75
						83	50	67	83	71
						100	33	33	67	58

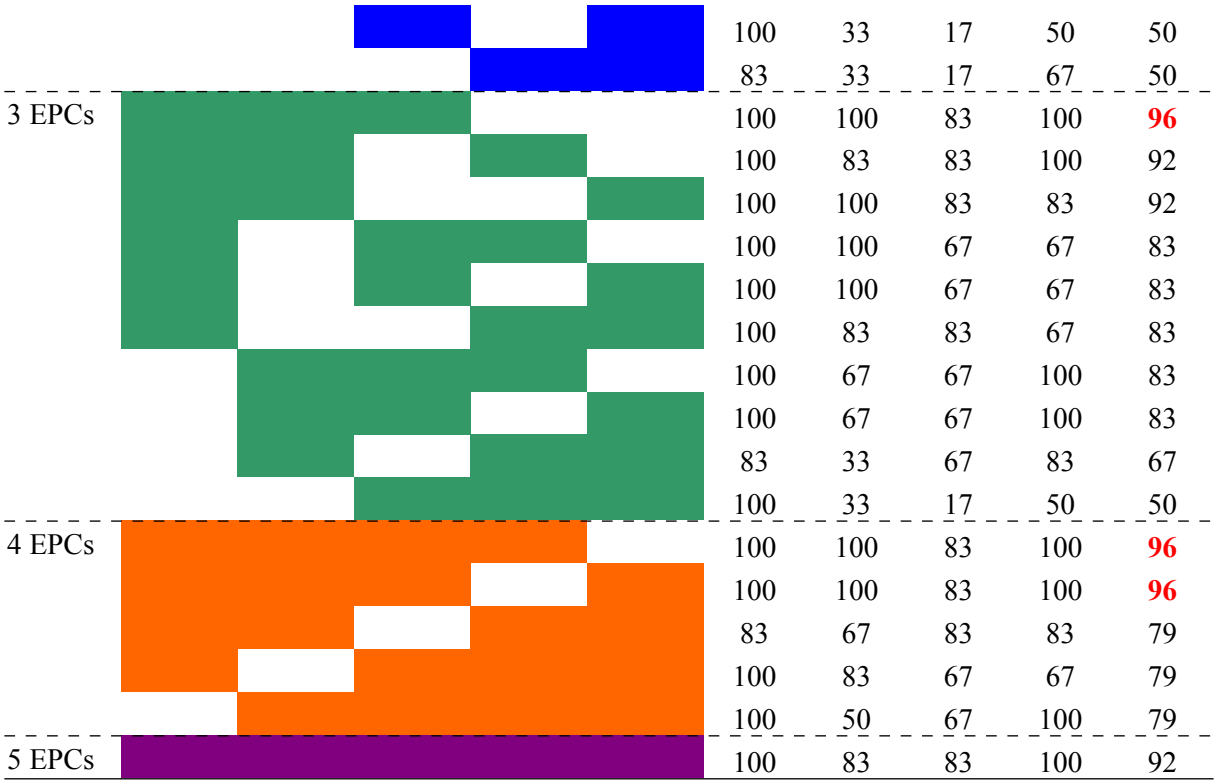

Table S5. Discrimination of 16 unknown homologous albumins by the PIC sensor array consisting of GAO/**P1**, GAO/**P2** and GAO/**P3**

Identification	Enzyme activity pattern			Verification	Accuracy
	GAO/**P1**	GAO/**P2**	GAO/**P3**		
HALB	0.662	1.223	1.283	EALB	No
HALB	0.351	0.969	0.866	HALB	Yes
HALB	1.127	0.935	0.836	EALB	No
HALB	0.522	0.909	1.266	HALB	Yes
RALB	1.019	1.615	1.783	RALB	Yes
RALB	0.988	1.400	1.531	RALB	Yes
RALB	1.277	1.254	1.533	RALB	Yes
RALB	1.006	1.788	1.611	RALB	Yes
BALB	2.582	2.784	2.804	BALB	Yes
BALB	2.555	2.578	2.731	BALB	Yes
BALB	2.804	2.534	2.695	BALB	Yes
BALB	2.308	2.733	2.793	BALB	Yes
EALB	0.558	1.039	0.753	EALB	Yes
EALB	0.333	0.571	0.882	BALB	No
EALB	0.797	0.764	1.025	EALB	Yes
EALB	0.597	1.211	0.880	EALB	Yes

References

[1] Y. Akiyama, A. Harada, Y. Nagasaki and K. Kataoka, *Macromolecules*, 2000, **33**, 5841.

[2] T. Ishii, H. Otsuka, K. Kataoka and Y. Nagasaki, *Langmuir*, 2004, **20**, 561.

[3] (a) V. Butun, S. P. Armes and N. C. Billingham, *Macromolecules*, 2001, **34**, 1148; (b) A. Tamura, M. Oishi and Y. Nagasaki, *J. Control. Release*, 2010, **146**, 378; (c) M. Tamura, S. Ichinohe, A. Tamura, Y. Ikeda and Y. Nagasaki, *Acta. Biomater.*, 2011, **7**, 3354.

[4] A. Tamura, M. Oishi and Y. Nagasaki, *Biomacromolecules*, 2009, **10**, 1818.

[5] B. Bjellqvist, G.J. Hughes, C. Pasquali, N. Paquet, F. Ravier, J. C. Sanchez, S. Frutiger and D. Hochstrasser, *Electrophoresis*, 1993, **14**, 1023.

[6] C. N. Pace, F. Vajdos, L. Fee, G. Grimsley and T. Gray, *Protein Sci.*, 1995, **4**, 2411.

[7] M. E. Lienqueo, A. Mahn and J. A. Asenjo, *J. Chromatogr. A*, **2002**, 978, 71.

[8] R. Fraczkiewicz and W. Braun, *J. Comput. Chem.,* **1998**, 19, 319.