

Ya Jin<sup>1,2,3</sup>  
 Qi Yuan<sup>1\*</sup>  
 Jun Zhang<sup>1\*</sup>  
 Takashi Manabe<sup>4</sup>  
 Wen Tan<sup>1,3</sup>

<sup>1</sup>School of Bioscience and Bioengineering, South China University of Technology, Guangzhou, P. R. China

<sup>2</sup>Guangdong Provincial Key Laboratory of Fermentation and Enzyme Engineering, South China University of Technology, Guangzhou, P. R. China

<sup>3</sup>Key Laboratory of Industrial Biotechnology of Guangdong Higher Education Institutes, South China University of Technology, Guangzhou, P. R. China

<sup>4</sup>Faculty of Science, Ehime University, Matsuyama, Japan

Received December 4, 2014

Revised February 16, 2015

Accepted May 7, 2015

## Research Article

# Proteomic analysis of cellular soluble proteins from human bronchial smooth muscle cells by combining nondenaturing micro 2DE and quantitative LC-MS/MS. 2. Similarity search between protein maps for the analysis of protein complexes

Human bronchial smooth muscle cell soluble proteins were analyzed by a combined method of nondenaturing micro 2DE, grid gel-cutting, and quantitative LC-MS/MS and a native protein map was prepared for each of the identified 4323 proteins [1]. A method to evaluate the degree of similarity between the protein maps was developed since we expected the proteins comprising a protein complex would be separated together under nondenaturing conditions. The following procedure was employed using Excel macros; (i) maps that have three or more squares with protein quantity data were selected (2328 maps), (ii) within each map, the quantity values of the squares were normalized setting the highest value to be 1.0, (iii) in comparing a map with another map, the smaller normalized quantity in two corresponding squares was taken and summed throughout the map to give an “overlap score,” (iv) each map was compared against all the 2328 maps and the largest overlap score, obtained when a map was compared with itself, was set to be 1.0 thus providing 2328 “overlap factors,” (v) step (iv) was repeated for all maps providing  $2328 \times 2328$  matrix of overlap factors. From the matrix, protein pairs that showed overlap factors above 0.65 from both protein sides were selected (431 protein pairs). Each protein pair was searched in a database (UniProtKB) on complex formation and 301 protein pairs, which comprise 35 protein complexes, were found to be documented. These results demonstrated that native protein maps and their similarity search would enable simultaneous analysis of multiple protein complexes in cells.

### Keywords:

Cellular proteins / Native protein map / Nondenaturing micro 2DE / Protein complex / Quantitative LC-MS/MS  
 DOI 10.1002/elps.201400574



Additional supporting information may be found in the online version of this article at the publisher's web-site

## 1 Introduction

Protein interactions in cells have been studied to reconstruct the complex structures and functions of cells. Yeast two-

hybrid system was employed for the global analysis of protein interaction networks in *E. coli* strain DY330 [2], *Saccharomyces cerevisiae* [3], and human [4]. Since this approach is limited to the detection of binary complexes and not directly aimed to analyze the protein interactions occurring in their physiological conditions, complementary information was obtained by employing affinity purification of targeted protein complexes and their analysis by MS [5]. Also, biochemical fractionation procedures were combined with mass spectrometric analysis for the analysis of protein complexes in cellular soluble

**Correspondence:** Professor Wen Tan, School of Bioscience and Bioengineering, South China University of Technology, Guangzhou, Guangdong 510006, P. R. China

**Email:** went@scut.edu.cn

**Fax:** +86-20-39380669

**Abbreviations:** CRTAP, cartilage-associated protein; HB-SMC, human bronchial smooth muscle cells; HDMS<sup>E</sup>, ion mobility separation-enhanced MS/MS in data-independent acquisition mode; nano-UPLC, nanoultra performance liquid chromatography

\*These authors contributed equally to this study.

**Colour Online:** See the article online to view Figs. 1–4 in colour.

fractions using ion-exchange HPLC, sucrose gradient centrifugation, and isoelectric focusing [6] and using blue native electrophoresis [7].

We have been employing nondenaturing micro 2DE for the analysis proteins and protein interactions in human plasma [8] and *E. coli* cytosol [9, 10]. In these works, CBB-stained protein spots on the 2DE gels were excised and MALDI-MS-PMF was used as a method of protein assignment. However, it was reported that even spots looking well separated on a 2DE gel could consist of several proteins [11]. Also, in the course of examining the performance of quantitative LC-MS/MS to analyze the proteins on nondenaturing 2DE gels, we realized that the sensitivity of the apparatus in protein detection exceeded that of conventional protein staining methods. So we developed the combined method of nondenaturing micro 2DE, grid gel-cutting, and quantitative LC-MS/MS that enabled not only the comprehensive analysis of proteins in the grid area, but also the reconstruction of quantity distribution maps (native protein maps) of all the identified proteins [12]. This method was applied for the analysis of HBSMC soluble proteins and 4323 proteins were identified in a 30 mm × 40 mm gel area providing the same number of native protein maps [1].

In this paper, we report on the comparisons of the native protein maps of HBSMC soluble proteins aiming at the acquisition of information on protein–protein interactions. Since each protein map was characterized by several features, such as the position of quantity peak square, number of detected squares, degree of concentration (focused or dispersed), etc., the similarity of two protein maps would suggest that the two proteins migrated together as a protein complex. Since it was difficult to visually compare the maps and judge the similarity between the 4323 protein maps, we developed Excel macros to extract protein pairs with similar maps and examined whether the protein pairs were described in the protein database UniProtKB.

## 2 Materials and methods

### 2.1 Materials, cell culture, protein sample preparation, and nondenaturing micro 2DE

The materials, the procedures of HBSMC culture, the preparation and the nondenaturing micro 2DE of the soluble protein fraction of HBSMC proteins were performed as previously described [1].

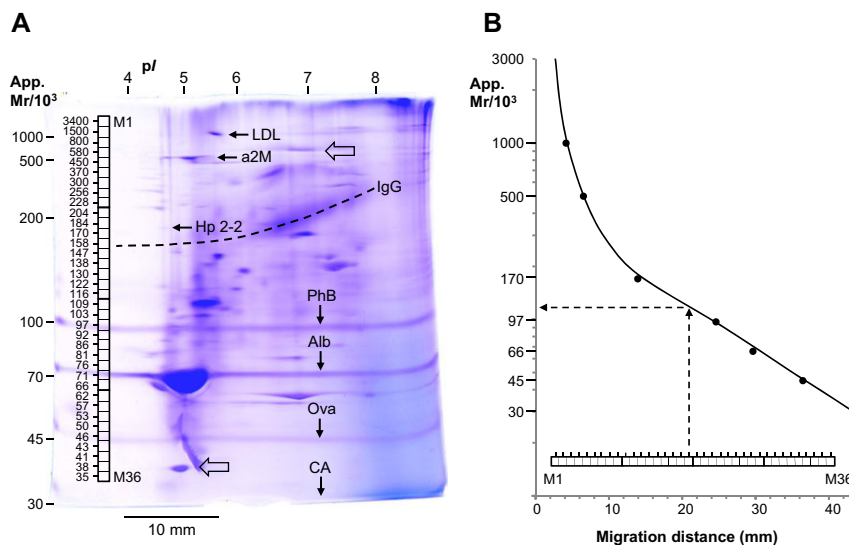
### 2.2 Coelectrophoresis of HBSMC soluble proteins with plasma proteins and LMW calibration proteins

The estimation of apparent *pI* and apparent mass of HBSMC proteins on the nondenaturing 2DE gel was done using the following procedure. A 10- $\mu$ L aliquot (*ca.* 160  $\mu$ g protein) of the fraction of HBSMC soluble protein, which contained 2 mM EDTA, 1 mM PMSF, and 12% w/v glycerol, was mixed

with 1  $\mu$ L human plasma (containing 40% w/v sucrose) and the mixture was applied on the top (basic end) of an agarose IEF micro column gel (1.4 mm id × 47 mm length) that contained 1% w/v agarose and 5% v/v of Pharmalyte pH 3–10 (20-fold dilution of the commercial solution). The catholyte was 0.04 M NaOH-0.01 M NaCl and the anolyte was 0.01 M phosphoric acid, both precooled in ice water. IEF was run at 0.12 mA/gel constant current until the voltage increased to 300 V (about 23 min) and then continued at 300 V constant voltage for 25 min. The IEF gel was extruded by water pressure onto a glass plate and the acidic end of the gel was cut off to give a final length of 37 mm, then the gel was transferred onto the top of a polyacrylamide micro slab gel (4.2–17.85% T linear gradient, 5% C, 42 mm high × 38 mm wide × 1 mm thick), where a 100- $\mu$ L aliquot of a 0.01 M Tris-0.02 M glycine buffer (pH 9.0) was added beforehand. A solution of low molecular weight (LMW) calibration proteins (GE Healthcare, Little Chalfont, UK) (0.576  $\mu$ g/ $\mu$ L), supplemented with 40% w/v sucrose, was applied at the both ends of the IEF gel (each 4  $\mu$ L) and on the IEF gel along the length of the gel (10  $\mu$ L) and the second-dimension electrophoresis was run using a 0.05 M Tris-0.10 M glycine buffer (pH 9.0) at 10 mA/gel constant current and stopped 45 min after the line of BPB migrated out of the gel bottom (37 min) (totally 82 min). The micro 2DE gel after CBB staining was shown in Fig. 1A.

### 2.3 In-gel digestion, nano-UPLC-MS/MS, and data processing

The gel pieces were subjected to the procedures of destaining, reduction, alkylation, and in-gel trypsin digestion, as described in detail in [12]. The extracted peptides were dried by vacuum centrifugation, reconstituted with a 10  $\mu$ L-aliquot of 1% v/v formic acid-2% v/v acetonitrile and supplemented with a 2  $\mu$ L-aliquot of 240 fmol/ $\mu$ L Enolase Digest Standard (Waters Corp., Milford, MA, USA) prepared in 1% v/v formic acid-2% v/v acetonitrile. Nano-UPLC (nano-ultra performance liquid chromatography) separation of the peptides was performed with a nanoACQUITY system equipped with a Symmetry 5  $\mu$ m C<sub>18</sub>, 180  $\mu$ m × 20 mm trap column and a UPLC 1.7  $\mu$ m BEH130 C<sub>18</sub>, 75  $\mu$ m × 100 mm analytical reverse phase column (all from Waters). On-line mass spectrometric measurement of the nano-UPLC-separated tryptic peptides was performed using a hybrid mass spectrometer coupling ion mobility separation with Q-TOF analyzer (Synapt G2-S HDMS, Waters) and equipped with a nano-ESI source. LC-MS/MS data were collected in resolution, positive and ion mobility separation-enhanced MS/MS in data-independent acquisition mode (HDMS<sup>E</sup>) mode, i.e. ion mobility separation-enhanced MS/MS in data-independent acquisition mode, using settings that have been optimized based on the manufacturer recommendations. The nano-LC-HDMS<sup>E</sup> data were processed with ProteinLynx Global SERVER (PLGS) *ver.* 2.5.2 (Waters). Data were lock mass calibrated post acquisition. Peak processing parameters were the low



**Figure 1.** The procedure to estimate apparent masses of HBSMC soluble proteins on a nondenaturing micro 2DE gel that was subjected to grid gel-cutting. (A) A nondenaturing micro 2DE gel obtained by coelectrophoresis of HBSMC soluble proteins and human plasma proteins in the step of IEF and further coelectrophoresis with LMW calibration proteins in the step of gradient gel electrophoresis, after CBB staining (details are described in Section 2.2). (B) A standard curve of apparent mass versus migration distance was prepared from the result in (A), using the following data points: human plasma low-density lipoprotein (LDL, 1000 kDa),  $\alpha$ 2-macroglobulin (a2M, 500 kDa), haptoglobin 2-2 polymer (Hp 2-2, the smallest polymer in the polymer series, 170 kDa), rabbit muscle phosphorylase b (PhB, 97 kDa), bovine serum albumin (Alb, 66 kDa), chicken egg ovalbumin (Ova, 45 kDa) and bovine erythrocyte carbonic anhydrase (CA, 30 kDa). Apparent masses of the 36 rows of the grid of the gel cutting, from M1 to M36 [1], were read out from (B) and indicated beside the squares in (A). The band shape of human plasma immunoglobulin G (IgG, 150 kDa), as indicated by the dotted line in (A), was used to estimate the degree of the retardation of basic proteins.

energy threshold 20 counts, high energy threshold ten counts and intensity threshold 500 counts. Database searching were performed using the following parameters: database, UniProtKB homo sapiens complete proteome dataset (canonical sequences only, 20 251 entries, 2013-05-29); peptide and fragment tolerance, both automatic (typically <10 ppm and <20 ppm, respectively [13]); maximum of missed trypsin cleavage, 1; maximum protein mass, 600 kDa; fixed modification, carbamidomethylation at Cys; variable modifications, oxidation at Met; false-positive rate (protein level), 4%. The criteria of protein identification were set as; at least two peptide matches per protein, at least three fragment ion matches per peptide, at least seven fragment ion matches per protein, and protein score above 100. Protein quantities were calculated by PLGS referring to the quantity of the internal standard (tryptic peptides of ENO1\_YEAST). Calculated masses of the proteins assigned by nano-LC-HDMS<sup>E</sup> were calculated using a laboratory-made Visual Basic program after searching the information of each protein in Protein Knowledgebase (UniProtKB) and getting the protein chain sequence (without signal peptide and propeptide). Further details of the procedure of nano-UPLC-MS/MS were described previously [1].

## 2.4 Preparation of protein maps of identified proteins

One of the aims of grid gel-cutting was to reconstruct the quantity distribution patterns (native protein maps) of the assigned proteins. The procedures previously used to

draw maps of  $5 \times 18$  squares for the analysis of human plasma high-density lipoprotein and its apolipoproteins [12] were extended to draw maps of  $27 \times 36$  squares for the analysis of HBSMC soluble proteins using Excel macros; (i) put a tag of the square number to all the data rows that include information on protein name, protein quantity, etc. (about 80 000 rows), (ii) collected all data in one worksheet and sorted them by “protein entry name” as the first priority and “quantity in ng” as the second priority to align each assigned protein in the order of quantity, (iii) copied the data of each assigned protein (maximum 967 lines with different square numbers) to a new worksheet, (iv) converted the values in the column of “quantity in ng” in each protein’s worksheet to percent values, setting the highest quantity to be 100%, making a new column of “percent quantity,” (v) drew  $27$  (wide)  $\times$   $36$  (high) squares to form a grid on each worksheet and to paint each square with a color, the transparency in % of the color was determined by calculating  $[100\% - \text{“percent quantity” of the square}]$ . Using the procedure, the distribution of a protein, in terms of its relative quantity within the grid area, was reconstructed as a color density pattern (a native protein map).

## 3 Results and discussion

### 3.1 Preparation of a standard curve for apparent mass estimation on protein maps

In order to examine the status of proteins detected on non-denaturing 2DE gels, whether they form protein complexes

or not, we used the information obtained by the comparisons between their calculated masses and apparent masses in the analysis of human plasma proteins [8] and *E. coli* soluble proteins [9]. So in the analysis of protein complexes in HBSMC soluble proteins, we prepared a standard curve to estimate apparent masses of proteins on the protein maps that have been prepared as described in Section 2.4. HBSMC soluble proteins were subjected to coelectrophoresis with plasma proteins and LMW calibration proteins as described in Section 2.2 and the micro 2DE gel after CBB staining was shown in Fig. 1A. The following molecular mass values were employed to prepare the standard curve of apparent mass (Fig. 1B); human plasma low-density lipoprotein, 1000 kDa; human plasma  $\alpha_2$ -macroglobulin (a2M), 500 kDa; human plasma haptoglobin phenotype 2-2 polymer (Hp 2-2, the smallest polymer in the polymer series), 170 kDa; rabbit muscle phosphorylase b (PhB), 97 kDa; bovine serum albumin (Alb), 66 kDa; chicken egg ovalbumin (Ova), 45 kDa; bovine erythrocyte carbonic anhydrase, 30 kDa. Then the positions of the 36 rows of the grid-cut gel area along the apparent mass direction (M1–M36) [1] were aligned on the pattern of Fig. 1A using the two protein spots indicated by the outline arrows. In order to facilitate to read out the apparent mass values on each protein map, the positions of M1–M36 were drawn on the standard curve, along the direction of migration distance in Fig. 1B, the apparent mass values at the center of the squares were calculated, and the values were shown beside the rows in Fig. 1A. Since we employed nondenaturing conditions in the second dimension run, it is expected that proteins that have *pI* values closer to the pH of the electrophoresis buffer would show smaller mobility. We tried to reduce the effects by using a buffer of pH 9.0 and extended run time of electrophoresis, then the band shape of human plasma IgG was used to estimate the degree of retardation of basic proteins (Fig. 1A dotted curve). These results were used in judging the possibility of interactions (binding) between proteins, as described in Section 3.4.

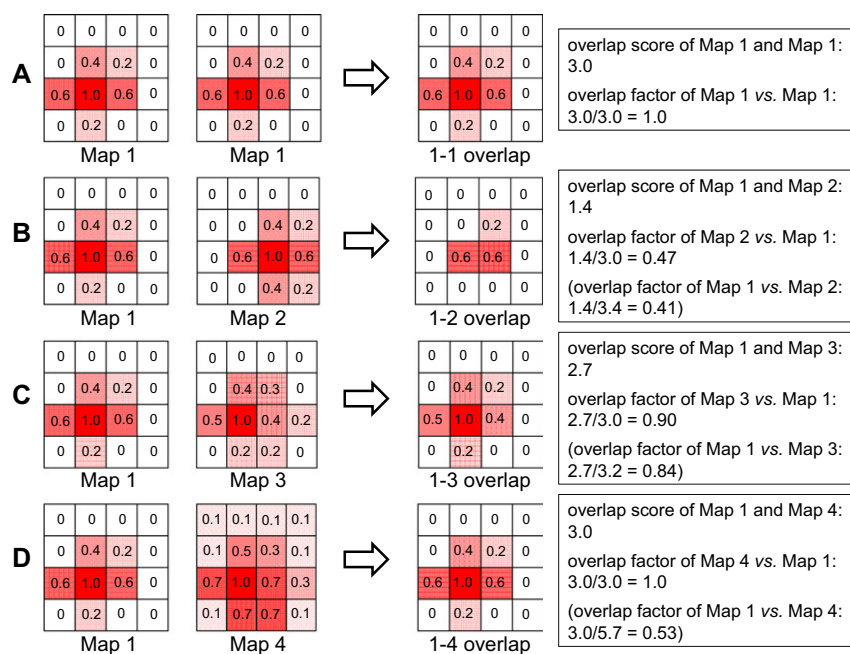
### 3.2 Evaluation of similarity between protein maps

We obtained native protein maps for the identified 4323 proteins and found each map can be differentiated from others by multiple features including the position of quantity peak in *pI* and apparent mass axes, number of squares in which the protein was detected, degree of concentration (focused or dispersed), shape of the detected square group (lengths in horizontal and vertical directions), etc. Therefore, if two protein maps were quite similar, it might suggest their binding throughout the process of 2DE since the electrophoretic separation was done under nondenaturing conditions. However, in order to examine the similarity between maps we needed to define the “degree of similarity” between a reference map and a sample map. Also, the process must be automated because the comparison would be done between thousands of maps, which means between more than one million map pairs. Therefore, we developed the following method to evaluate the degree of similarity between the maps and to select

protein pairs that show similar maps, using Excel macros written in Visual Basic for Applications (VBA). (1) The proteins that have been detected in 3 or more squares within  $27 \times 36$  squares were selected (2328 maps). (2) The quantity values of the squares in a map were normalized setting the highest quantity value to be 1.0. (3) When a map was compared with another map, the smaller normalized quantity from the two corresponding squares was taken and the sum throughout the map was designated as an “overlap score.” (4) Each map was compared against the 2328 protein maps and the largest overlap score, obtained when a map was compared with itself, was set to be 1.0 thus providing 2328 “overlap factors” against each map. (5) The step (4) was repeated for all maps providing  $2328 \times 2328$  matrix of overlap factors and the protein pairs that showed an overlap factor above a threshold value from both protein sides were selected. Figure 2 explains the steps (2) and (3) of the method described above, using simplified model maps of  $4 \times 4$  squares. As shown in Fig. 2A, when Map 1 is set as a reference and compared with itself, the overlap score and the overlap factor can be calculated to be 3.0 and 1.0, respectively. When Map 2 that has a density peak position different from Map 1 is compared with Map 1 (Fig. 2B), an overlap score of 1.4 and an overlap factor ( $1.4/3.0 =$ ) 0.47 of Map 2 against Map 1 are obtained. Map 3, which has a density peak at the same position as Map 1, provides a high overlap factor of 0.90. The comparison between a pair of maps should be done setting each one as a reference map, as typically shown in Fig. 2D. When a protein showed wide distribution and the quantity peak position was not clear, a map like Map 4 is obtained and it provides overlap factor of 1.0 against Map 1. However, when Map 4 is set as a reference and compared with Map 1, a much smaller overlap factor of 0.53 is obtained. We tentatively named the method to search similar protein maps as “overlap search.”

### 3.3 Application of overlap search to 2328 protein maps

Figure 3 illustrates the steps (4) and (5) of the method described in Section 3.2, using proteasome subunit alpha type 1 (PSA1\_HUMAN) as an example. When the map of PSA1 was set as a reference and compared with the 2328 maps, overlap factors were obtained as shown in Fig. 3. The area around PSA1 showed proteins with high overlap scores against PSA1 and the area was expanded as Fig. 3. The overlap factor of PSA1 was 1.0 because it was compared with itself, and there were a series of proteins that showed high overlap factors (PSA2–PSA7 and PSB1–PSB8) in this area. The comparisons were repeated setting each of the 2328 maps and the values of overlap factors comprised a  $2328 \times 2328$  matrix, in which the overlap factors shown in Fig. 3A formed one of the rows. Figure 3C shows an overlap factor matrix around PSA1, specially prepared to illustrate the concept of the  $2328 \times 2328$  matrix. It visualized the presence of many protein pairs with a high overlap factor (a square densely colored) from both sides of the proteins, such as PSA2 against PSA1 (the value at row “PSA1” and column “PSA2”) and PSA1



**Figure 2.** The concept of “overlap score” and overlap factor” developed for objective selection of similar protein maps. Simplified model maps of  $4 \times 4$  squares were used to illustrate the steps of the calculation of “overlap score” and “overlap factor.” (A) When Map 1 is set as a reference and compared with itself, overlap score and overlap factor can be calculated to be 3.0 and 1.0, respectively. (B) When Map 2 that has a density peak position different from Map 1 is compared with Map 1, an overlap score of 1.4 and an overlap factor of  $(1.4/3.0 = )$  0.47 are obtained. (C) Map 3 that has a density peak at the same position as Map 1 provides a high overlap factor of 0.90. (D) When a protein showed wide distribution and the quantity peak position was not as clear as in Map 4, it provides an overlap factor of 1.0 against Map 1. However, when Map 4 is set as a reference and compared with Map 1, a much smaller overlap factor of 0.53 is obtained. Therefore, the comparison between a pair of maps should be done setting each one as a reference map.

against PSA2 (the value at row “PSA2” and column “PSA1”). The maps of 16 proteins, PSA1 to PSA7 and PSB1 to PSB9, were shown in Fig. 3D in order to correlate the matrix with actual protein maps. Each protein in the group of 15 proteins, PSA1–PSA7 and PSB1–PSB8, formed a protein pair with at least one of the other 14 proteins with high overlap factors. However, PSB9 did not show high overlap factors when each of the 15 proteins was set as a reference (Fig. 3C, the values at column “PSB9”), although the map of PSB9 suggested that the protein might comprise a complex together with the 15 proteins. The results of low overlap factors might be caused by the relatively lower quantity of PSB9 compared with others (about 1/90 to 1/5), which was around the detection limit of the apparatus for each square, then the quantity distribution map was not drawn correctly.

Fourteen proteins, PSA1–PSA7 and PSB1–PSB7, showed a similar quantity level as shown in Fig. 3D. Since these proteins have similar calculated masses, these results suggested that they might form an equimolar complex. In fact, search in UniProtKB provided the description that these 14 proteins comprise the 20S proteasome core with  $(14 \times 2 =)$  28 subunits that are arranged in four stacked rings, resulting in a barrel-shaped structure [14]. Also, it was described that PSB5 and PSB6 can be replaced by PSB8 and PSB9, respectively, which explained the relatively lower quantities of these two compared with the other 14 proteins. All of the 16 proteins showed apparent mass values about

450 kDa at their distribution center and the mass of 20S proteasome core could be calculated from the calculated masses of the subunits to be 669.2 kDa. The discrepancy might be explained that the subunits are compactly stacked to apparently show a smaller mass value than calculated. In our results, the regulatory subunits of proteasome showed map patterns different from the core subunits (Section 3.4). We speculate that 26S proteasomes were decomposed into two parts in the course of cell disruption (sonication).

### 3.4 Selection of similar protein map pairs and database search

As described in Section 3.3, we prepared a  $2328 \times 2328$  matrix of overlap factors and we could decide any threshold value of overlap factors to choose similar protein map pairs. When the threshold values were set at 0.50, 0.60, 0.65, 0.70, 0.75, and 0.80, the number of selected protein pairs were 1904, 724, 431, 241, 132, and 56, respectively. It was expected that the higher the threshold value the higher the ratio of reported protein complexes within the selected protein pairs. However, since we aimed to evaluate the method of overlap search not only to confirm reported protein complexes but also to predict possible protein complexes, a threshold of overlap factor 0.65 from both sides of the proteins, which we judged would cover most of the candidate protein pairs, was

**Table 1.** Summary of the search of complex formation in database UniProtKB on the 431 protein pairs that showed overlap factor above 0.65 for each other

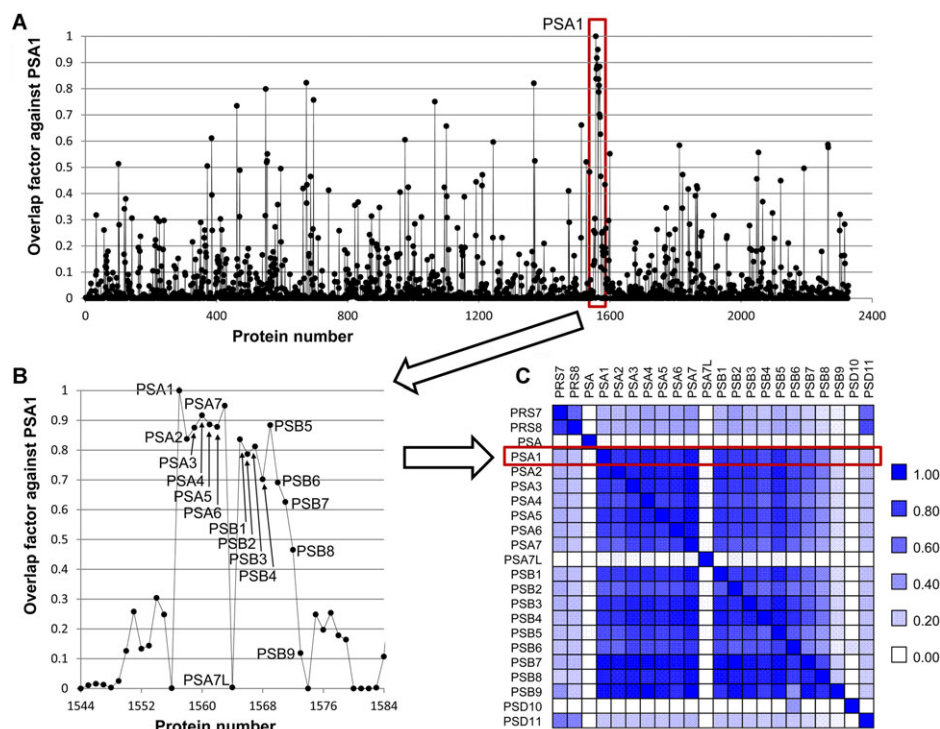
Protein complex number	Entry names of searched protein couple(s) <sup>a)</sup>	Name of protein complex in UniProtKB	File number in Supporting Information 2 <sup>b)</sup>	Reference number <sup>c)</sup>
1	PSA1, PSA2, PSA3, PSA4, PSA5, PSA6, PSA7, PSB1, PSB2, PSB3, PSB4, PSB5, PSB6, PSB7, PSB8	26S proteasome 20S core	File10, File51,File105	[14]
2	ADRM1, ECM29, PRS4, PRS6A, PRS6B, PRS8, PRS10, PSD7,PSD12, PSD13, PSDE, PSMD1, PSMD3, PSMD6, PSMD8, UCHL5	26S proteasome / 19S regulatory subunit	File107	[14]
3	ICT1, RM15	39S ribosome mitochondria	File81	[16]
4	AP2A1, AP2M1	Adaptor protein complex 2 (AP-2)	File17	[17]
5	ARL2, TBCD	ARL2-TBCD complex	File19	[18]
6	ARP2, ARP3, ARPC2, ARPC3, ARPC4	Arp2/3 complex	File20	[19]
7	BRE1A, BRE1B	BRE1 complex	File24	[20]
8	CAN2, CPNS1	Calpain	File25	[21]
9	COPA, COPB, COPB2, COPG1, COPZ1	Coatomer complex	File39	[22]
10	COMD5, COMD7, COMDA, DSCR3	COMM domain protein complex	File38	[23]
11	CND1, CND2	Condensin complex	File37	[24]
12	CSN1, CSN2, CSN3, CSN5, CSN6, CSN7A, CSN8	COP9 signalosome complex	File41	[25]
13	XRCC5, XRCC6	DNA-dependent protein kinase complex DNA-PK	File138	[26]
14	EF1B, EF1G	EF-1 complex	File53	[27]
15	VPS25, VPS36	Endosomal sorting complex required for transport II (ESCRT-II)	File86	[28]
16	(STAM2, STAM1, HGS)	ESCRT-0 complex	File44	[29]
17	IF2A, IF2B, IF2G	Eukaryotic translation initiation factor 2	File82	[30]
18	EIF3A, EIF3E, EIF3F	Eukaryotic translation initiation factor 3 (eIF-3) complex	File57	[31]
19	MGN, RBM8A	Exon junction complex (EJC)	File91	[32]
20	FRIH, FRIL	Ferritin 24 mer	File66	[33]
21	RRAGA, RRAGC	Rag complex	File124	[34]
22	ETFA, ETFB	ETFA and ETFB	File60	[35]
23	PDIA3, PDIA6	Large shaperon complex	File98	[36]
24	MCM6, MCM7	MCM complex	File90	[37]
25	DDX3X, DHX9	mRNP complex	File46	[38]
26	SYEP, SYIC	Multisynthetase complex	File131	[39]
27	UBA3, ULA1	NEDD8-activating enzyme E1	File134	[40]
28	CRTAP, P3H1	Newly proposed in human <sup>d)</sup>	File40	[15]
29	NU205, NUP93	Nuclear pore complex (NPC)	File95	[41]
30	HDAC2, MTA2	Nucleosome remodeling and histone deacetylation (NuRD) complex	File76	[42]
31	VATD, VATE1, VATG1, VATH	Peripheral V1 complex of vacuolar ATPase	File137	[43]
32	RL3, RL4, RL6, RL7, RL7A, RL8, RL9, RL11, RL13, RL14,RL15, RL18, RL19, RL26, RL27A, RL28, RL29, RL31,RL32, RL35, RL36L, RS2, RS3A, RS4X, RS6, RS9, RS10,	80S ribosome	File113 - File122, File125 - File127	[44]
32	RS13, RS14, RS15A, RS16, RS18, RS19, RS23, RS25	80S ribosome	File113 - File122, File125 - File127	[44]
33	POP1, RPP30	RNase P	File101	[45]
34	SEP11, SEPT2, SEPT7	Septin complex	File129	[46]
35	TCPA, TCPB, TCPD, TCPE, TCPG, TCPH, TCPQ, TCPZ	T-complex protein 1	File133	[47]
36	PRP8, U520	U4/U6-U5 tri-snRNP complex	File104	[48]

a) The entry names (without the part "\_HUMAN") of the protein couples searched in UniProtKB. When two or more couples were described to compose one protein complex, their entry names were combined in one cell.

b) Details of the database search results were summarized as files as provided in Supporting Information 2.

c) One paper that described the subunit structure of the corresponding protein complex was cited.

d) This complex was described in the database for chicken, but not in human.



**Figure 3.** Explanation of the method to find similar protein maps using the steps described in Section 3.3. (A) The protein map of proteasome subunit alpha type 1 (PSA1\_HUMAN) was set as a reference map and overlap factors against 2328 protein maps that had three or more detected squares were calculated and plotted. The x coordinates are the numbers of the 2328 proteins sorted by names in alphabetic order. (B) The area around PSA1 in (A) was expanded and the entry names of proteins were shown. The overlap factor of PSA1 was 1.0 because it was compared against itself. (C) The calculation results like in (A) were accumulated setting each of the 2328 maps forming a  $2328 \times 2328$  matrix of overlap factors, but here only a matrix around PSA1 was shown to illustrate the concept of the overlap factor matrix. For simplicity, the value of overlap factor in each square was replaced with a color density. (D) (on next page) In order to correlate the matrix in (C) with actual protein maps, the maps of 16 proteins, PSA1–PSA7 and PSB1–PSB9, were shown. Each map was added with the UniProt protein entry name (without “.HUMAN”), number of squares detected, percent abundance against the total protein quantity within the grid area, and mass calculated from the amino acid sequence. Details are described in Section 3.3.

tentatively decided. The number of protein pairs above the threshold 0.65 was 431 and they were selected out of  $((2328 \times 2328 - 2328)/2 = 2708628)$  protein pairs. Each of the protein in the 431 protein pairs were examined in the sections “Function” and “Interaction/Subunit structure” of the database UniProtKB and we found 301 protein pairs were described to form protein complexes. These results strongly suggested that the overlap search of native protein maps would be useful in visualizing and confirming the presence of protein complexes. Also, the selected threshold value was proved to be appropriate to cover the candidate protein maps pairs. The results of the search in UniProtKB were summarized in Supporting Information 1 and 2. The 301 protein pairs were reported to comprise 35 human protein complexes, as summarized in Table 1. Then we further examined the remaining 130 protein pairs in STRING, a database of known and predicted protein interactions (<http://string-db.org/>), in order to expand the search from confirmed description of binding in UniProtKB to predicted description of binding in STRING. Although the number of human proteins covered in STRING was smaller than UniProtKB, STRING has the search function “multiple names” that was convenient to search known and

predicted interactions between a pair of proteins. Within 130 protein pairs, only one pair of proteins, cartilage-associated protein (CRTAP) and prolyl 3-hydroxylase 1 (P3H1), showed a high score in STRING search on the possibility of binding and a paper that reported a three-protein complex between the two proteins and cyclophilin B (peptidyl-prolyl cis-trans isomerase B, PPIB) in chicken [15] was found. Then we reexamined the maps of the three proteins and found PPIB was also detected at the positions of CRTAP and P3H1, but PPIB also distributed at a basic and high-apparent mass region that resulted in the overlap factors under the threshold against CRTAP and P3H1. This complex was also shown in Table 1, because we judged that the presence of the three-protein complex in HBSMC is highly probable (see also File40 in Supporting Information 2). The positions of the protein complexes listed in Table 1 were illustrated in Fig. 4. Some of the complexes were not shown in Fig. 4 since it was difficult to directly illustrate their distribution patterns, and details of all the protein complexes were given in Supporting Information 1 and 2. The 129 protein pairs without description on complex formation were further examined on the possibility of complex formation, mainly by the comparisons of their calculated

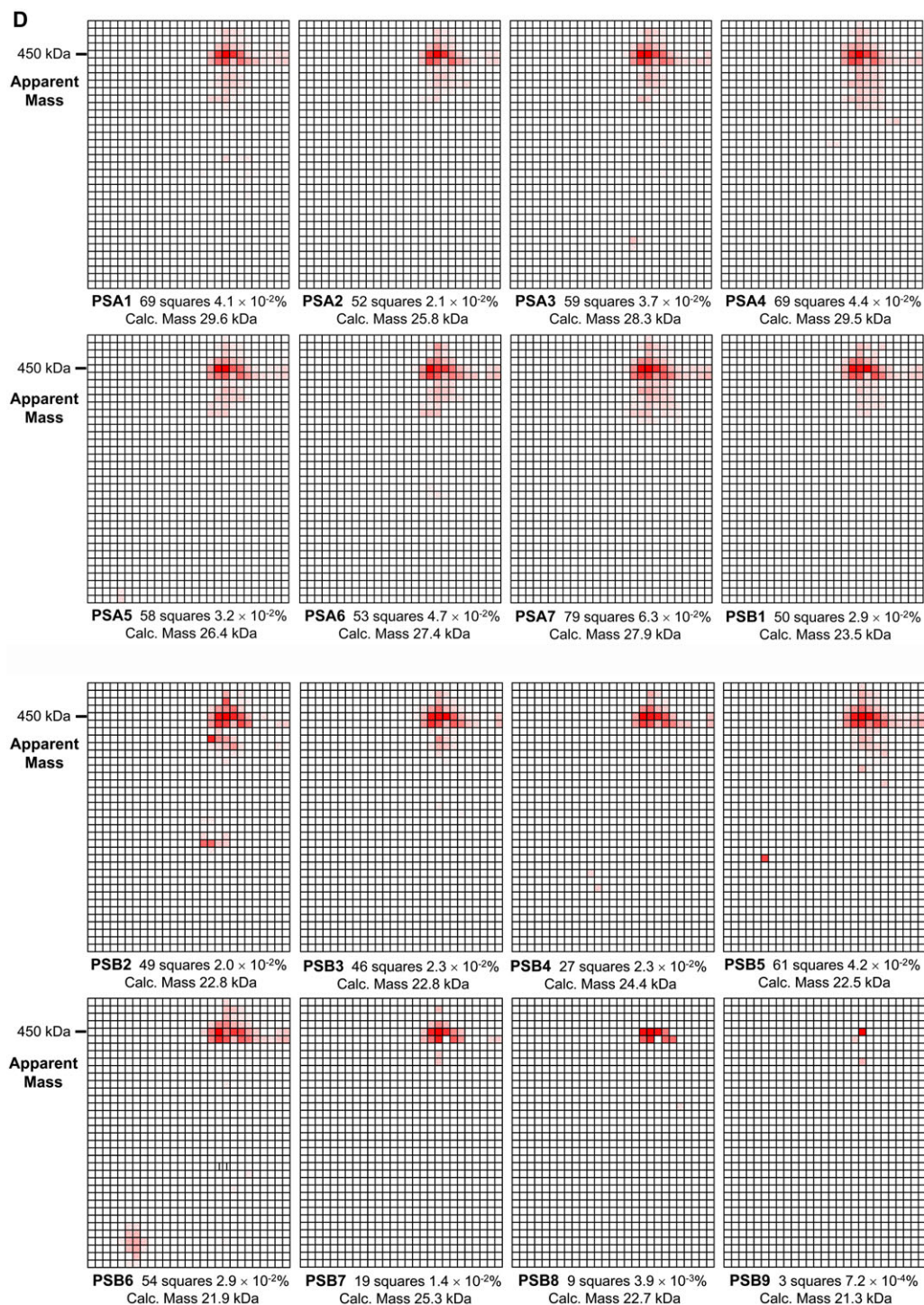
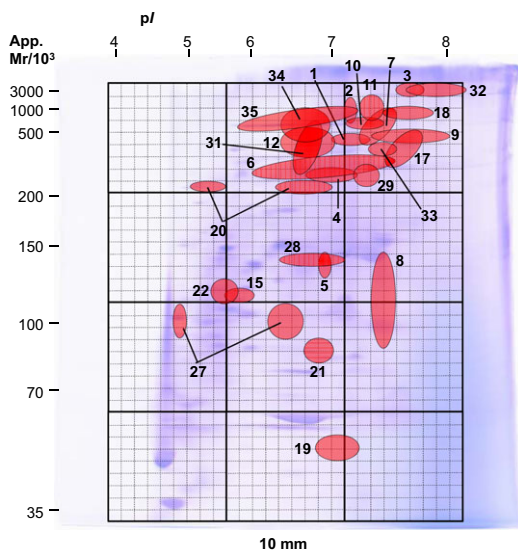


Figure 3. Continued.

masses and apparent masses, since a protein complex would show an apparent mass value close to the sum of the calculated masses of the two proteins. We tentatively proposed that 50 protein pairs out of the 129 pairs do not form complexes,

but as for the remaining 79 pairs we could not get enough information in UniProtKB to judge the presence or absence of their binding. The results of protein map comparisons, database search, and the comparisons of apparent mass and





**Figure 4.** The positions of the Table 1-listed protein complexes on the nondenaturing 2D gel. The numbers correspond to the number of the protein complexes in Table 1 (Column 1). 1, 26S proteasome 20S core; 2, 26S proteasome regulatory subunit; 3, 39S ribosome mitochondria; 4, adaptor protein complex 2 (AP-2); 5, ARL2-TBCD complex; 6, Arp2/3 complex; 7, BRE1 complex; (8, Calpain); 9, coatomer complex; 10, COMM domain protein complex; 11, condensin complex; 12, COP9 signalosome complex; (13, DNA-dependent protein kinase complex DNA-PK); (14, EF-1 complex); 15, endosomal sorting complex required for transport II (ESCRT-II); (16, ESCRT-0 complex); 17, eukaryotic translation initiation factor 2; 18, eukaryotic translation initiation factor 3; 19, exon junction complex; 20, ferritin 24 mer; (21, Rag complex); 22, ETFA and ETFB; (23, large shaperon complex); (24, MCM complex); (25, mRNP complex); (26, multisynthetase complex); 27, NEDD8-activating enzyme E1; 28, CRTAP-P3H1-PPIB complex (newly proposed in human); 29, nuclear pore complex (NPC); (30, nucleosome remodeling and histone deacetylation (NuRD) complex); 31, Peripheral V1 complex of vacuolar ATPase; 32, 80S ribosome, 33, RNase P; 34, septin complex; 35, T-complex protein 1; (36, U4/U6-U5 tri-snRNP complex). The complexes in the parentheses are not shown in the figure for their complex distribution patterns. Details on all the protein complexes were given in Supporting Information 1 and 2.

calculated mass, were summarized in a file for each protein pair or protein group as Supporting Information 2. Although each protein map provided information on the state of the protein as homodimer or homooligomer by comparing its apparent mass value with its calculated mass, we focused on the method of overlap search between multiple protein maps and the information was not summarized in this paper. These results will be described in detail elsewhere.

### 3.5 Maps of low-abundant proteins

As described in Section 3.2, we used 2328 protein maps that had three or more detected squares for the overlap search of protein maps, because we judged that the concept of overlap factor would not efficiently work on the comparisons of maps

with one or two detected squares. When 1531 protein maps with one detected square were compared with each other on the overlapping of the detected square, 85 maps overlapped at 13 different squares that provided 350 protein pairs. The search option of “multiple search” in STRING was applied to each group of proteins detected in one of the 13 squares, and 15 protein pairs were reported by STRING to show high scores on protein interactions. These results showed that the protein maps with only one detected square also contain information useful in analyzing protein interactions. However, the method of protein map comparison described in Section 3.2–3.4 applied to the maps with three or more detected squares provided more concrete information on multiple protein complexes. Further improvements in the sensitivity of protein identification, such as the use of a Fourier transform ion cyclotron resonance mass spectrometer or an Orbitrap mass spectrometer, would enable to detect the low-abundant proteins in larger numbers of squares and expand the applicability of the overlap search method to all 4323 proteins or more.

## 4 Concluding remarks

We obtained 4323 native protein maps of HBSMC soluble proteins using a combined method of nondenaturing micro 2DE, grid gel-cutting, and quantitative LC-MS/MS. A method to evaluate the degree of similarity between protein maps with three or more detected squares (2328 maps) was developed introducing the concept of “overlap factor” and a matrix of overlap factors (2328 × 2328) was prepared to select protein map pairs with high similarity. Out of the selected 431 pairs, 301 protein pairs were found to be documented in a database UniProtKB to form protein complexes to comprise 35 protein complexes. These results demonstrated that the overlap search method described here enabled simultaneous analysis of multiple cellular protein complexes on nondenaturing 2D gels. Improvements of the sensitivity of LC-MS/MS system would further expand the applicability of the method in searching cellular protein complexes present in low quantities.

*The mass spectrometry proteomic data have been deposited to the ProteomeXchange Consortium (<http://www.proteomeexchange.org>) via the PRIDE partner repository [49] with the dataset identifier PXD001471. The authors would like to thank the PRIDE team for their great help with the data deposition process. This work is financially supported by “National Natural Science Foundation of China 31400798,” “The Fundamental Research Funds for the Central Universities 2013ZM0048,” “Science and Technology Program of Guangzhou, China 2014J4100022,” “Guangdong Province Key Laboratory of Fermentation and Enzyme Engineering Open Fund, China FJ2013002,” and “Ministry of Science and Technology of Guangdong Province, China 2012A080204001.”*

*The authors have declared no conflict of interest.*

## 5 References

- [1] Jin, Y., Zhang, J., Yuan, Q., Manabe, T., Tan, W., *Electrophoresis* 2015, in press (DOI 10.1002/elps.201400573).
- [2] Butland, G., Peregrín-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., Davey, M., Parkinson, J., Greenblatt, J., Emili, A., *Nature* 2005, 433, 531–537.
- [3] Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A. P., Punna, T., Peregrín-Alvarez, J. M., Shales, M., Zhang, X., Davey, M., Robinson, M. D., Paccanaro, A., Bray, J. E., Sheung, A., Beattie, B., Richards, D. P., Canadien, V., Lalev, A., Mena, F., Wong, P., Starostine, A., Canete, M. M., Vlasblom, J., Wu, S., Orsi, C., Collins, S. R., Chandran, S., Haw, R., Rilstone, J. J., Gandi, K., Thompson, N. J., Musso, G., St Onge, P., Ghanny, S., Lam, M. H., Butland, G., Altaf-Ul, A. M., Kanaya, S., Shilatifard, A., O'Shea, E., Weissman, J. S., Ingles, C. J., Hughes, T. R., Parkinson, J., Gerstein, M., Wodak, S. J., Emili, A., Greenblatt, J. F., *Nature* 2006, 440, 637–643.
- [4] Venkatesan, K., Rual, J. F., Vazquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Zenkner, M., Xin, X., Goh, K. I., Yildirim, M. A., Simonis, N., Heinzmann, K., Gebreab, F., Sahalie, J. M., Cevik, S., Simon, C., de Smet, A. S., Dann, E., Smolyar, A., Vinayagam, A., Yu, H., Szeto, D., Borick, H., Dricot, A., Klitgord, N., Murray, R. R., Lin, C., Lalowski, M., Timm, J., Rau, K., Boone, C., Braun, P., Cusick, M. E., Roth, F. P., Hill, D. E., Tavernier, J., Wanker, E. E., Barabási, A. L., Vidal, M., *Nat. Methods* 2009, 6, 83–90.
- [5] Ewing, R. M., Chu, P., Elisma, F., Li, H., Taylor, P., Climie, S., McBroom-Cerajewski, L., Robinson, M. D., O'Connor, L., Li, M., Taylor, R., Dharsee, M., Ho, Y., Heilbut, A., Moore, L., Zhang, S., Ornatsky, O., Bukhman, Y. V., Ethier, M., Sheng, Y., Vasilescu, J., Abu-Farha, M., Lambert, J. P., Duetzel, H. S., Stewart, I. I., Kuehl, B., Hogue, K., Colwill, K., Gladwish, K., Muskat, B., Kinach, R., Adams, S. L., Moran, M. F., Morin, G. B., Topaloglou, T., Figeys, D., *Mol. Syst. Biol.*, 2007, 3, 89.
- [6] Havugimana, P. C., Hart, G. T., Nepusz, T., Yang, H., Turinsky, A. L., Li, Z., Wang, P. I., Boutz, D. R., Fong, V., Phanse, S., Babu, M., Craig, S. A., Hu, P., Wan, C., Vlasblom, J., Dar, V. U., Bezginov, A., Clark, G. W., Wu, G. C., Wodak, S. J., Tillier, E. R., Paccanaro, A., Marcotte, E. M., Emili, A., *Cell* 2012, 150, 1068–1081.
- [7] Wessels, H. J., Vogel, R. O., Lightowlers, R. N., Spelbrink, J. N., Rodenburg, R. J., van den Heuvel, L. P., van Gool, A. J., Gloerich, J., Smeitink, J. A., Nijtmans, L. G., *PLoS One* 2013, 8, e68340.
- [8] Manabe, T., Jin, Y., Tani, O., *Electrophoresis* 2007, 28, 843–863.
- [9] Manabe, T., Jin, Y. *Electrophoresis* 2010, 31, 2740–2748.
- [10] Manabe, T., Jin, Y., *Electrophoresis* 2011, 32, 300–309.
- [11] Schmidt, F., Schmid, M., Jungblut, P. R., Mattow, J., Facius, A., Pleissner, K.-P., *J. Am. Soc. Mass Spectrom.* 2003, 14, 943–956.
- [12] Jin, Y., Bu, S., Zhang, J., Yuan, Q., Manabe, T., Tan, W., *Electrophoresis* 2014, 35, 2055–2064.
- [13] Li, G. Z., Vissers, J. P., Silva, J. C., Golick, D., Gorenstein, M. V., Geromanos, S. J., *Proteomics* 2009, 9, 1696–1719.
- [14] Wang, X., Chen, C.-F., Baker, P. R., Chen, P.-L., Kaiser, P., Huang, L., *Biochemistry* 2007, 46, 3553–3565.
- [15] Ishikawa, Y., Wirz, J., Vranka, J. A., Nagata, K., Bächinger, H. P., *J. Biol. Chem.* 2009, 284, 17641–17647.
- [16] Richter, R., Rorbach, J., Pajak, A., Smith, P. M., Wessels, H. J., Huynen, M. A., Smeitink, J. A., Lightowlers, R. N., Chrzanowska-Lightowlers, Z. M., *EMBO J.* 2010, 29, 1116–1125.
- [17] Pearse, B. M., Smith, C. J., Owen, D. J., *Curr. Opin. Struct. Biol.* 2000, 10, 220–228.
- [18] Bhamidipati, A., Lewis, S. A., Cowan, N. J., *J. Cell Biol.* 2000, 149, 1087–1096.
- [19] Welch, M. D., Iwamatsu, A., Mitchison, T. J., *Nature* 1997, 385, 265–269.
- [20] Zhu, B., Zheng, Y., Pham, A.-D., Mandal, S. S., Erdjument-Bromage, H., Tempst, P., Reinberg, D., *Mol. Cell* 2005, 20, 601–611.
- [21] Strobl, S., Fernandez-Catalan, C., Braun, M., Huber, R., Masumoto, H., Nakagawa, K., Irie, A., Sorimachi, H., Bourenkow, G., Bartunik, H., Suzuki, K., Bode, W., *Proc. Natl. Acad. Sci. USA* 2000, 97, 588–592.
- [22] Lowe, M., Kreis, T. E., *J. Biol. Chem.* 1996, 271, 30725–30730.
- [23] Burstein, E., Hoberg, J. E., Wilkinson, A. S., Rumble, J. M., Csomos, R. A., Komarck, C. M., Maine, G. N., Wilkinson, J. C., Mayo, M. W., Duckett, C. S., *J. Biol. Chem.* 2005, 280, 22222–22232.
- [24] Schmiesing, J. A., Gregson, H. C., Zhou, S., Yokomori, K., *Mol. Cell. Biol.* 2000, 20, 6996–7006.
- [25] Fang, L., Wang, X., Yamoah, K., Chen, P. L., Pan, Z. Q., Huang, L., *J. Proteome Res.* 2008, 7, 4914–4925.
- [26] Walker, J. R., Corpina, R. A., Goldberg, J., *Nature* 2001, 412, 607–614.
- [27] Sheu, G. T., Traugh, J. A., *J. Biol. Chem.* 1997, 272, 33290–33297.
- [28] Langelier, C., von Schwedler, U. K., Fisher, R. D., De Domenico, I., White, P. L., Hill, C. P., Kaplan, J., Ward, D., Sundquist, W. I., *J. Virol.* 2006, 80, 9465–9480.
- [29] Bache, K. G., Raiborg, C., Mehlum, A., Stenmark, H., *J. Biol. Chem.* 2003, 278, 12513–12521.
- [30] Paytubi, S., Morrice, N. A., Boudeau, J., Proud, C. G., *Biochem. J.* 2008, 409, 223–231.
- [31] Zhou, M., Sandercock, A. M., Fraser, C. S., Ridlova, G., Stephens, E., Schenauer, M. R., Yokoi-Fong, T., Barsky, D., Leary, J. A., Hershey, J. W. B., Doudna, J. A., Robinson, C. V., *Proc. Natl. Acad. Sci. USA* 2008, 105, 18139–18144.
- [32] Zhao, X. F., Nowak, N. J., Shows, T. B., Aplan, P. D., *Genomics* 2000, 63, 145–148.
- [33] Arosio, P., Adelman, T. G., Drysdale, J. W., *J. Biol. Chem.* 1978, 253, 4451–4458.
- [34] Sekiguchi, T., Hirose, E., Nakashima, N., Ii, M., Nishimoto, T., *J. Biol. Chem.* 2001, 276, 7246–7257.
- [35] McKean, M. C., Beckmann, J. D., Frerman, F. E., *J. Biol. Chem.* 1983, 258, 1866–1870.
- [36] Meunier, L., Usherwood, Y.-K., Chung, K. T., Hendershot, L. M., *Mol. Biol. Cell* 2002, 13, 4456–4469.
- [37] Ishimi, Y., *J. Biol. Chem.* 1997, 272, 24508–24513.

- [38] Weidensdorfer, D., Stoehr, N., Baude, A., Lederer, M., Koehn, M., Schierhorn, A., Buchmeier, S., Wahle, E., Huettelmaier, S. *RNA* 2009, *15*, 104–115.
- [39] Mirande, M., Le Corre, D., Louvard, D., Reggio, H., Pailiez, J. P., Waller, J. P., *Exp. Cell Res.* 1985, *156*, 91–102.
- [40] Bohnsack, R. N., Haas, A. L., *J. Biol. Chem.* 2003, *278*, 26823–26830.
- [41] Krull, S., Thyberg, J., Bjorkroth, B., Rackwitz, H. R., Cordes, V. C., *Mol. Biol. Cell* 2004, *15*, 4261–4277.
- [42] Schmidt, D. R., Schreiber, S. L., *Biochemistry* 1999, *38*, 14711–14717.
- [43] Adachi, I., Arai, H., Pimental, R., Forgac, M., *J. Biol. Chem.* 1990, *265*, 960–966.
- [44] Anger, A. M., Armache, J. P., Berninghausen, O., Habeck, M., Subklewe, M., Wilson, D. N., Beckmann, R., *Nature* 2013, *497*, 80–85.
- [45] Jarrous, N., Eder, P. S., Wesolowski, D., Altman, S., *RNA* 1999, *5*, 153–157.
- [46] Nagata, K., Asano, T., Nozawa, Y., Inagaki, M., *J. Biol. Chem.* 2004, *279*, 55895–55904.
- [47] Rommelaere, H., Van Troys, M., Gao, Y., Melki, R., Cowan, N. J., Vandekerckhove, J., Ampe, C., *Proc. Natl. Acad. Sci. USA* 1993, *90*, 11975–11979.
- [48] Liu, S., Rauhut, R., Vornlocher, H. -P., Luehrmann, R., *RNA* 2006, *12*, 1418–1430.
- [49] Vizcaíno, J. A., Deutsch, E. W., Wang, R., Csordas, A., Reisinger, F., Ríos, D., Dianes, J. A., Sun, Z., Farrah, T., Bandeira, N., Binz, P. A., Xenarios, I., Eisenacher, M., Mayer, G., Gatto, L., Campos, A., Chalkley, R. J., Kraus, H. J., Albar, J. P., Martinez-Bartolomé, S., Apweiler, R., Omenn, G. S., Martens, L., Jones, A. R., Hermjakob, H., *Nat. Biotechnol.* 2014, *30*, 223–226.