A Nonredundant Human Protein Chip for Antibody Screening and Serum Profiling*

Angelika Lueking‡, Alexandra Possling‡, Otmar Huber§, Allan Beveridge‡, Martin Horn¶, Holger Eickhoff¶, Johannes Schuchardt‖, Hans Lehrach‡, and Dolores J. Cahill‡‡‡**

There is burgeoning interest in protein microarrays, but a source of thousands of nonredundant, purified proteins was not previously available. Here we show a glass chip containing 2413 nonredundant purified human fusion proteins on a polymer surface, where densities up to 1600 proteins/cm^2 on a microscope slide can be realized. In addition, the polymer coating of the glass slide enables screening of protein interactions under nondenaturing conditions. Such screenings require only 200-μl sample volumes, illustrating their potential for high-throughput applications. Here we demonstrate two applications: the characterization of antibody binding, specificity, and cross-reactivity; and profiling the antibody repertoire in body fluids, such as serum from patients with autoimmune diseases. For the first application, we have incubated these protein chips with anti-RGSHis$_6$, anti-GAPDH, and anti-HSP90β antibodies. In an initial proof of principle study for the second application, we have screened serum from alopecia and arthritis patients. With analysis of large sample numbers, identification of disease-associated proteins to generate novel diagnostic markers may be possible. ***Molecular & Cellular Proteomics 2:1342–1349, 2003.***

Array-based technologies are increasing in significance in genomic and proteomic research. DNA and oligonucleotide microarrays are widely used tools (reviewed in Refs. 1–3). Currently, one of the main applications of DNA chips is for gene expression profiling, a key technology in genomics to unravel differential gene expression (4–7). Although DNA array technology has been used very successfully for a variety of applications, and protein chips would complement DNA chips as a logical development, the current use of protein chips is limited to only a small number of laboratories (reviewed in Refs. 8 and 9). To a large extent this is due to the more complex nature of proteins, requiring laborious work in cloning, recombinant expression, and purification. Here, we have addressed this problem by developing an automated, robust protein expression and purification protocol to provide thousands of highly pure human proteins and used them to generate high-density, high-content protein arrays on polymer-coated glass slides, with up to 1600 proteins per cm^2. Current methods to determine antibody binding and specificity are still labor-intensive and complicated. They involve screening the antibodies on nonordered, λgt11 phage library plaque-lift filters or on tissue extracts previously separated by two-dimensional gel electrophoresis (10–12).

Applying protein microarray technology will greatly simplify this process, especially the protein arrays described here contain large numbers of proteins derived from an ordered recombinant source. This enables the direct connection from the protein product of the individual expression clone to its corresponding cDNA sequence information (13). Other advantages include the short time and small screening volumes required. Moreover, many experiments can be performed in parallel, and the entire procedure is suitable for automation.

Büssow *et al.* have previously shown that cDNA libraries, cloned into an *Escherichia coli* expression vector, can be screened for protein expression on high-density filter membranes (14). Using robot technology (15), a human fetal brain cDNA expression library (hEx1) was picked into microtitre plates, and high-density protein arrays were produced on filter membranes followed by *in situ* protein expression. This approach was extended to automated spotting of protein microarrays from crude lysates of the expression cultures (16). In an initial analysis of this library, more than 66% of the clones contain inserts in the correct reading frame. Sixty-four percent of these clones comprise full-length proteins. Furthermore, these expressed proteins are suitable for either matrix-assisted laser desorption/ionisation-time-of-flight mass spectrometry and functional screening assays (17, 18). This library has been normalized to obtain a nonredundant, unique cDNA set (Uniclone Set) using oligonucleotide fingerprinting (19). Automated, parallel protein expression and purification were performed on 2413 protein-expressing clones, as described in "Experimental Procedures." Purified proteins were immobilized on modified glass surfaces. The protein microarrays

From the ‡Max-Planck-Institute for Molecular Genetics, Ihnestrasse 73, 14195 Berlin, Germany; §Institute of Clinical Chemistry and Pathobiochemistry, Universitätsklinikum Benjamin Franklin, Hindenburgdamm 30, 12200 Berlin, Germany; ¶Scienion AG, Volmerstrasse 7a, D-12489 Berlin, Germany; ‖MicroDiscovery GmbH, Immanuelkirchstraβe 12, D-10405 Berlin, Germany; and ‡‡Centre for Human Proteomics, Royal College of Surgeons in Ireland, Dublin 2, Ireland and Protagen AG, Emil-Figge-Straβe 76a, D-44227 Dortmund, Germany

Received, March 10, 2003, and in revised form, August 25, 2003
Published, MCP Papers in Press, September 29, 2003, DOI 10.1074/mcp.T3000010-MCP200

generated were tested for suitability to determine antibody specificity, and their applicability for profiling of serum from patients with autoimmune diseases was analyzed.

EXPERIMENTAL PROCEDURES

Expression Vector and Bacterial Strain—A cDNA library (hEx1) was generated from human fetal brain poly(A)$^+$ RNA by oligo(dT) priming, was size fractionated by gel filtration and directionally (SalI-NotI) cloned in the expression vector pQE30NST (GenBank Accession No. AF074376) and transformed into *Escherichia coli* strain SCS1 (Stratagene, La Jolla, CA) (14). The protein expression subset of this library was normalized by oligo-fingerprinting and re-arrayed into a nonredundant subset. A total of 2413 clones of this nonredundant set were selected and used for high-throughput expression.

Sequence and Data Analysis of hEX1 cDNA Expression Library—A total of 2303 clones of the expression library were sequenced, and the sequences obtained were translated in the three possible reading frames (+1, +2, +3) by the bioinformatics package GCG (Wisconsin Package Version 10.3, Accelrys Inc., San Diego, CA). All translated amino acid sequences were stringently searched for the epitope tag sequence, RGSHHHHHH (RGSH$_6$-tag), and the final reading frame of the expression clone was determined.

Subsequently, the 2303 cDNA sequences were blasted against the nonredundant protein database, NRPROT. The BLAST results for each cDNA clone were searched, and the top five hits were tabulated. The final reading frame determined by the RGSH$_6$-tag was determined, and the results were analyzed by searching for BLAST hits, which match the correct reading frame. Four sets of results were generated:

1. sequences that had at least one top-five BLAST hit that corresponded to the correct reading frame;
2. sequences that produced BLAST hits that did not correspond to the correct reading frame;
3. sequences that generated BLAST hits, but where the correct reading frame is unknown; and
4. sequences that did not match anything in NRPROT.

Protein Expression and Purification in High-throughput—Each protein was expressed twice in 1-ml cultures in deep-well microtitre plates. The protein extracts of each culture were generated as described (16) and combined. Proteins were purified as described with the modification, that the proteins were eluted in 30 μl 35% (v/v) acetonitrile, 0.1% (v/v) trifluoroacetic acid (17).

Generation of Protein Chips—A microscope slide treated with Bind-Silane (Amersham Pharmacia Biotech, Piscataway, NJ) and a covering slide treated with Repel-Silane (Amersham Pharmacia Biotech) were arranged together, separated by a 30-μm thick metal spacer. The space between the slides was filled with solution of 8% (w/v) polyacrylamide-bisacrylamide (30:0.8). After polymerization, the slides were washed with water and dried. The microscope slides were placed in a Q-Array System (Genetix, New Milton, UK), equipped with humidity control (50%) and 16 blunt-ended stainless steel print tips with a tip diameter of 150 μm. A 13 $\times$ 13 spot pattern with a center-to-center distance of 420 μm was printed onto the slides as duplicates. Each spot was loaded five times, resulting in a total transfer volume of 10 nl. These ready-to-use slides can be used without loss of signal giving reproducible results for at least 4 weeks when stored at 4 °C.

Detection Procedure—After spotting, the protein chips were blocked in 2% (w/v) bovine serum albumin/Tris-buffered saline, 0.1% (v/v) Tween 20 (TBST)[1] at room temperature and incubate with primary antibody (mouse-anti-RGSH$_6$ (Qiagen, Valencia, CA) 1:2000 dilution; mouse-anti-GAPDH (Research Diagnostics, Inc., Flanders, NJ) clone 6C5, 1:5,000 dilution; mouse-anti-HSP90β (Transduction Laboratories, Lexington, KY) clone 68, 1:2,000 dilution), followed by two 10-min TBST washes and incubation with the secondary antibody (rabbit-anti-mouse-IgG-Cye3 (Dianova, Hamburg, Germany), 1:800 dilution) in 2% (v/v) bovine serum albumin/TBST. Subsequently, the protein chips were washed three times for 10 min in TBST-T (0.5% (v/v) Triton X-100) followed by a 5-min wash with Tris-buffered saline. Detection of the signals obtained on these protein microarrays was performed using a 428^{TM} Arrayscanner System (Affimetrix, Palo Alto, CA). Protein microarrays used for serum profiling were blocked in 5% (w/v) fish gelatin/TBST at room temperature, and the serum was added (diluted 1:20 in 5% (w/v) fish gelatin/TBST). After two 10-min TBST washes and subsequent incubation with the secondary antibody (mouse-anti-human immunoglobulin G (IgG; Sigma, St. Louis, MO) 1:5,000 dilution) in 5% (w/v) fish gelatin/TBST, the protein chips were washed three times for 10 min in TBST. This was followed by incubation with the tertiary antibody (rabbit-anti-mouse-IgG-Cye3 (Dianova) 1:800 dilution) in 5% (w/v) fish gelatin/TBST. Subsequently, the arrays were washed three times, each in TBST-T for 20 min. Detection was performed as previously described. All incubation steps were performed for 1 h. All antibody dilutions were in blocking buffer unless otherwise stated.

Western Blot—For Western blot analysis, proteins were purified from 35-ml bacterial cultures. Protein expression was induced with 1 mM isopropyl β-D-thiogalactoside at an OD_{578} of 0.6. After 4 h cells were pelleted and resuspended in 3 ml 8 M urea, 0.1 M NaH_2PO_4, 10 mM Tris/HCl pH 8.0. After centrifugation, 1 ml of the cell lysate was purified with spin columns according to the manufactures instructions. Equal amounts of protein were separated by SDS-PAGE and transferred to polyvinylidene difluoride membrane by semi-dry blotting. The membrane was blocked in TBST for 1 h, incubated with patient serum diluted 1:500 in TBST for 1 h, and washed three times in TBST. Goat-anti-human IgG-peroxidase (Dianova) diluted 1:10,000 in TBST was added for 30 min, and after washing chemoluminescent detection was performed.

Image Analysis—Image and data analysis was performed using proprietary software developed by MicroDiscovery (Berlin, Germany). The clear and distinct signals and homogeneous spot shape obtained (as seen in Fig. 1*A*) facilitated image analysis. Data normalization was achieved by robust normalization techniques described in (20) and (21).

RESULTS

For the generation of protein chips, we analyzed the protein expressing resource, a nonredundant subset of a human fetal brain cDNA expression library. Sequence analysis of 2303 of the 2413 clones was performed. The reading frame of the cDNA insert was stringently determined with respect to the amino-terminal M-RGSH$_6$-tag. In parallel, the 2303 cDNA sequences were compared against the nonredundant protein database NRPROT, and the top five BLAST hits for each cDNA clone were found. The final reading frame was determined by the RGSH$_6$-tag, and the results were analyzed by searching for BLAST hits, which match the correct reading frame. The sequence matching criteria used were highly stringent in that there had to be a perfect match of all 27 nucleotides of the RGSH$_6$-tag. If there was a sequence error in this region, the sequence would be called as not having the histidine tag. Using these stringent criteria, no reading frame was

[1] The abbreviations used are: TBST, Tris-buffered saline, Tween-20; IgG, immunoglobulin G.

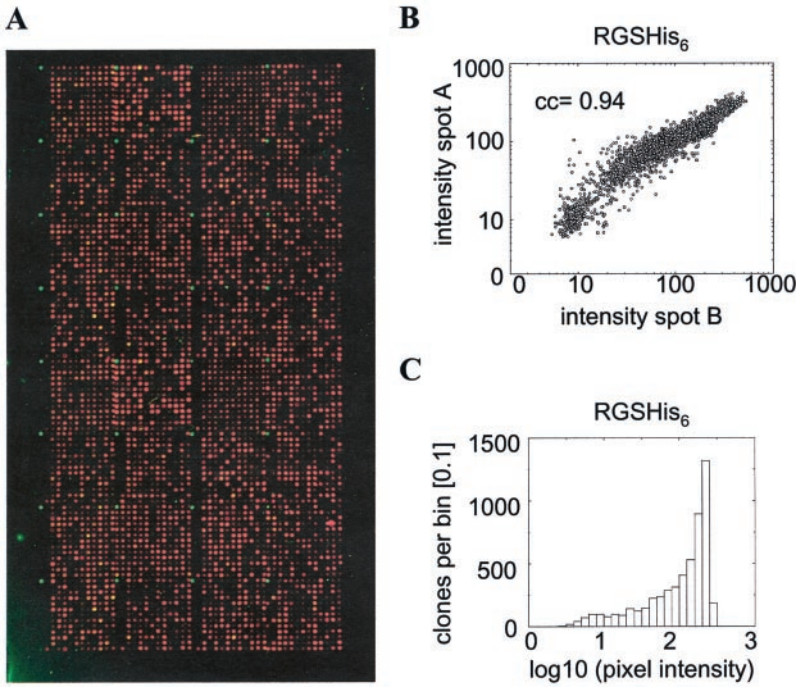

FIG. 1. **Protein microarray of 2413 recombinant, purified human proteins.** *A*, protein microarray probed with anti-$RGSH_6$ antibody. All proteins positively detected by the anti-$RGSH_6$ antibody are represented in *red* and the guide dots in *green*. *B*, correlation of relative spot intensities of duplicates (*spot A versus* its duplicate *spot B*). *C*, distribution of relative spot intensities.

determined for 701 clones. The remaining 1602 sequences gave the following results: 880 (55%) clones have their cDNA insert in the $RGSH_6$-tag corresponding reading frame, including 382 (43%) clones, comprising full-length proteins, and 724 (45%) clones that show a different reading frame to the frame determined by the $RGSH_6$-tag. Further analysis of the 701 clones, where the reading frame was not determined, showed that 616 of these clones had no intact $RGSH_6$-tag; 453 of these clones give a BLAST result, whereas 248 clones show no match at all.

To generate protein chips, proteins were expressed and purified in parallel under denaturing conditions from 2413 bacterial clones obtained from a human fetal brain cDNA expression library (14) that had been normalized by oligonucleotide fingerprinting (19). We have chosen *E. coli* as the expression system and optimized the culture volume and purification conditions leading to expression of suitable amounts for over 96.5% of the selected recombinant proteins (Fig. 1*C*). Following high-throughput protein expression and purification by nickel-nitrilotriacetic acid-immobilized metal affinity chromatography (22), the expressed proteins were spotted onto a modified glass surface, which has been coated with a thin polyacrylamide layer. The 2413 purified, human recombinant proteins were spotted in duplicate onto a glass microscope slide (25 × 75 mm) using a transfer stamp consisting of 16 pins with a 150-μm tip size. The duplicates were ordered into two areas, each area consisting of a 13 × 13 pattern (quadrants), corresponding to a total number of 5408 spots, including 32 guide dots, on one slide. Rabbit-anti-rat-IgG labeled with Cye5 served as a useful guide dot. Transferring 10 nl of each of the protein samples results in spots of about 220 μm in diameter, which enables theoretical spot densities of up to 1600 spots/cm^2. Due to the nonviscous and noncrystalline eluting reagent (35% (v/v) acetonitrile, 0.1% (v/v) trifluoroacetic acid), the protein spots, when loaded five times, gave a sharp, round, and well-localized signal. In contrast to DNA microarrays, ghost spots (*i.e.* spots with signal intensities below background) or doughnut effects with a non-Gaussian spot intensity profile were not observed on this surface.

To monitor the protein expression, purification, and arraying procedures, the protein chips were screened with an antibody specific for the $RGSH_6$-tag epitope of these His_6-tagged fusion proteins (Fig. 1*A*). When relative intensities of duplicates were plotted against each other, the resulting diagonal indicates a good reproducibility of the spotting and detection of the immobilized proteins with a correlation coefficient of 0.94 (Fig. 1*B*). Relative intensities of detected signals were rather homogeneously distributed in the range of 0–300 units with a peak value at 225 units (Fig. 1*C*). The average detected signal intensity was 132.5 units. We set a threshold of signals below 20 units as background; using this conservative threshold only 3.5% of the proteins were not detected.

When a high-speed picoliter-spotting robot (ink-jet device) was used for the deposition of proteins onto the polyacrylamide glass surface, the spot density could be increased to 5000 spots/cm^2. Spotting different concentrations of purified GAPDH assessed the sensitivity of detection of specific proteins (Fig. 2*A*). The sensitivity of detection was calculated to be 10 pmol/μl, corresponding 7.2 fmol or 288 pg of GAPDH in a volume of 720 pl spotted in four drops. When the proteins were spotted in a single step, a volume of 180 pl, with a minimal concentration of 40 pmol/μl, was required. GAPDH

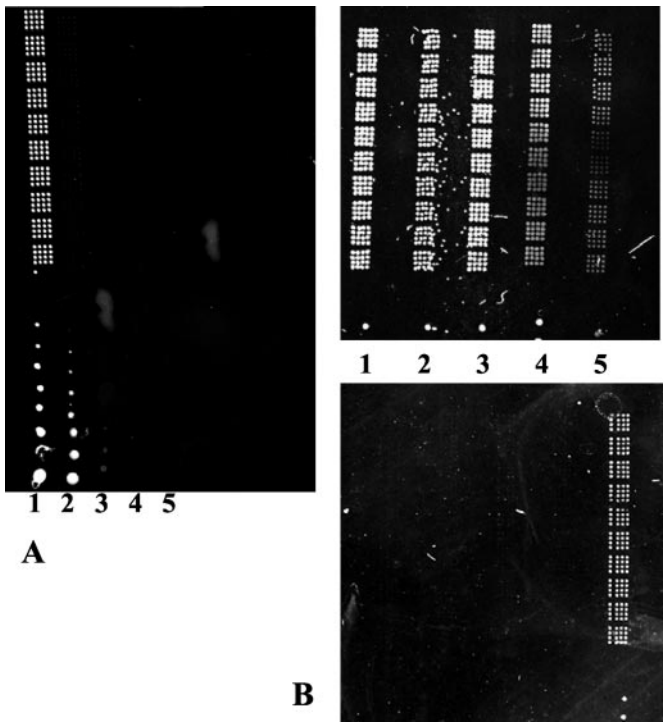

FIG. 2. **High-speed picolitre spotting (ink-jetting) of recombinant, purified proteins.** *A*, determination of the detection limit. Purified GAPDH was spotted at 100 pmol/μl, 10 pmol/μl, 1 pmol/μl, 100 fmol/μl, and 10 fmol/μl and were subsequently detected by anti-GAPDH antibody. *B*, detection of proteins purified in high-throughput using the epitope specific anti-$RGSH_6$ antibody (*top*) and a specific anti-GAPDH antibody (*bottom*). *Lanes 1–4*, unknown proteins; *lane 5*, GAPDH.

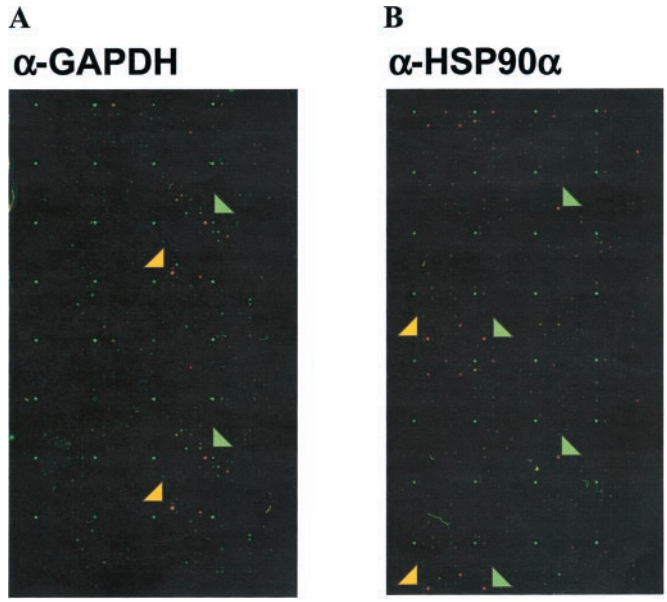

FIG. 3. **Determination of the specificity of antibodies against GAPDH (*A*) and HSP90β (*B*).** *Yellow arrows* indicate the corresponding antigens (*red spots*); *green arrows* indicate clones showing cross-reactivity with medium intensity. The guide dots appear as *green spots*.

protein and purified proteins from four other uncharacterized clones were piezoelectric sprayed onto the microchip and detected either by anti-$RGSH_6$ and anti-GAPDH antibody (Fig. 2, *B* and *C*). These results demonstrate that our high-throughput protein expression and purification procedure leads to sufficient protein concentrations to enable the reproducible generation of high-density protein microarrays.

As mentioned above, here we demonstrate the first of the two protein chip applications to be described, the characterization of antibody binding, specificity, and cross-reactivity. Antibodies are commonly used in diagnostic assays in many different formats (23), requiring a high assay sensitivity and specificity. We have analyzed and compared the specificity of two mouse monoclonal antibodies, anti-GAPDH and anti-HSP90β, respectively. Both antibodies are designed for their use in Western immunoblotting and therefore they recognize linear epitopes. Because the secondary mouse antibodies that were used to detect the monoclonal antibodies (anti-GAPDH and anti-HSP90β) might give some unspecific background signal, control incubations exclusively with the secondary antibody were performed. Clones, which were detected in both the control as well as the anti-GAPDH and anti-HSP90β incubations, were discarded. Moreover, the detailed sequence analysis of these clones has shown that these interactions were nonspecific.

The monoclonal anti-GAPDH antibody recognized its specific antigen GAPDH with the highest intensity of 305 units (616F11; Fig. 3*A*). Additionally, clone 613F03 coding for ribosomal protein S20 (RPS20) was recognized with a relative intensity of 201 units, respectively. The other weak signals detected had an intensity of one-fifth to one-tenth of the GAPDH signal. Ninety-nine percent of all clones show a signal less than the set threshold of 20 units. In the anti-HSP90β antibody screen, its target antigen (clone 616P24; relative intensity, 277 units; Fig. 3*B*) was preferentially detected, but in addition this antibody showed medium cross-reactivity with at least two other clones (with relative intensity of 151 and 149 units, respectively) and weak cross-reactivity to a further seven clones. The sequence analysis of the medium-intensity clones (homology to clone FLJ22122 fis, clone HEP19214 (no known function) and Gallus gallus nuclear calmodulin-binding protein) recognized by anti-HSP90β does not reveal any identical linear sequence motif. The proteins corresponding to the weak-intensity clones were similarly analyzed, but showed no homology to HSP90β.

Here, we demonstrate the second protein chip application, profiling the antibody repertoire in serum. Screening protein arrays with sera from large numbers of autoimmune patients, for example with rheumatoid arthritis, diabetes, multiple sclerosis, or alopecia, would not only allow the identification of potentially new autoantigens, but also the diagnosis and subtyping of the autoimmune disease based on the presence of specific autoantibodies (24). In this proof of principle serum

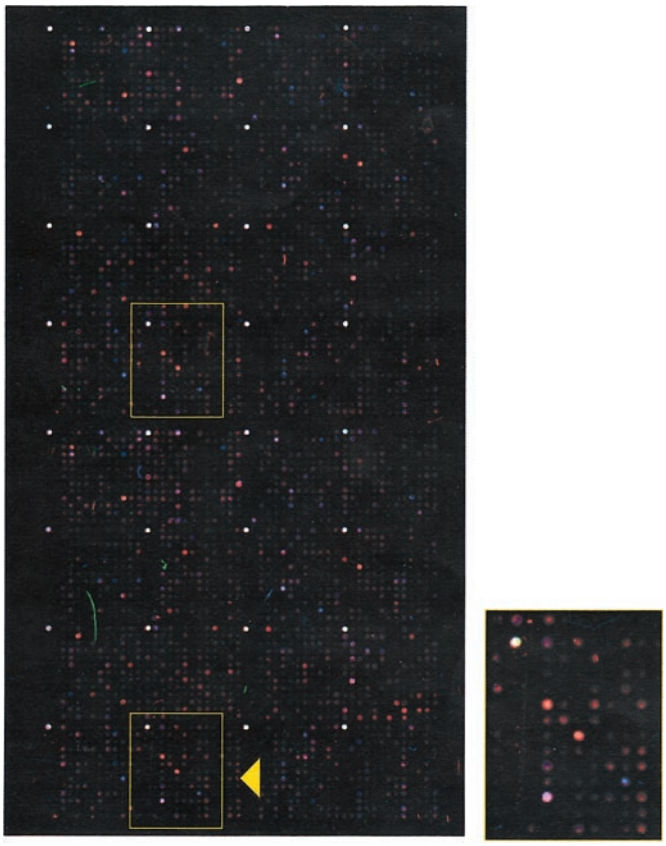

FIG. 4. **Serum profiling on the protein microarrays.** This composed image is the result of an overlay of an image from a protein microarray incubated with the control anti-human-IgG antibody (*blue spots*) with the image from a separate microarray incubated with the serum of a diffuse alopecia patient (*red spots*). The guide dots appear as *white spots*. Individual proteins recognized by both the control and the patient sera appear as *purple spots*.

profiling experiment, we screened sera from two patients with rheumatoid arthritis and one serum from a patient showing characteristic symptoms of diffuse alopecia. The datasets obtained from patient sera incubations were compared with control sera from clinically nonremarkable persons and to background incubations with human anti-IgG. From these results, proteins detected by the autoimmune sera were identified (Fig. 4). 5′-tag sequences of these clones were determined, and their sequences with respect to an open reading frame were analyzed. Three clones with false reading frames had to be excluded. The remaining sequences were used for BLAST searches against the public databases including GenBank and Unigene (25) (Table I). The relative intensities of the clones were compared for each serum screened on the protein microarray. Higher relative intensities for sera from diffuse alopecia and arthritis were obtained for the gene coding for autoantigen p69. The two sera of patients with rheumatoid arthritis recognized clones coding for a protein of the RAS association domain family 1 and for diglycerol kinase ζ. The alopecia serum recognized two additional proteins, the tumor suppressor p33 ING homologue and a protein coded by cDNA FLJ20427 with no known function. The microarray data were confirmed by analyzing equal amounts of the purified proteins (Fig. 5) on Western immunoblots with the diffuse alopecia patient serum. The control protein ubiquitin B, included in the Western immunoblot, was not detected by any serum. Controls with a serum obtained from an alopecia areata patient, another hair loss disease with clearly different clinical features, and with the second antibody alone were also screened (Fig. 5). The diffuse alopecia serum detected the autoantigen p69 and the tumor suppressor p33 ING homologous protein, whereas the gene product of cDNA FLJ20427 is not detected on the Western blot (Fig. 5*A*). This might be due to changes in the conformation after SDS treatment or to effects of the immobilization either on the chip or during Western immunoblot procedure. Additionally, weaker signals were obtained for the homologue of mouse synaptotagmin VI and the protein belonging to the RAS association domain family I. Both of these proteins were also recognized by the serum of the patient with characteristics specific for alopecia areata, whereas autoantigen p69 and tumor suppressor p33 ING homologous protein were not detected, suggesting different types of hair loss (Fig. 5*B*).

DISCUSSION

Here, a new method for the generation of protein microarrays on coated glass slides with a high sensitivity for antibody screening and serum profiling has been developed. Combining cDNA expression libraries and robot technology, 2413 different human proteins have been expressed, purified, and spotted onto glass chips.

Similar to DNA arrays, glass slides for protein chips offer a low fluorescent background, superior handling properties, require small screening volumes, and are suitable for automation. Recently, a microarray consisting of 10,800 spots of two single/distinct proteins (protein G an FKBP12 binding domain) has been described (26). Brown and coworkers have analyzed 115 antibody-antigen pairs arrayed on poly-L-lysine-coated glass slides and determined detection limits in the range of 100 pg/ml, confirming previously specified detection levels (16, 27). Because microarrays allow highly parallel analysis, there is increasing interest in arraying a large number of different targets. We have chosen *E. coli* as a robust, fast, and efficient protein expression system. Because the proteins were expressed in *E. coli*, no post-translational modifications such as phosphorylation or glycosylation were present. However, if protein expression systems introducing post-translational modifications such as *Saccharomyces cerevisiae*, *Pichia pastoris*, or baculovirus (28–31) would have been used, the glycosylation pattern obtained would not be identical to humans. We have shown that 96.5% of the 2413 clones express a detectable His_6-tagged fusion protein with an average signal intensity of 132.5 units. The sequence analysis of clones expressing proteins with a relative intensity close to

TABLE I

Candidate autoantigens detected by sera antibodies from three patients

The table includes the list of proteins identified when the protein chip was screened with serum from three patients, two with rheumatoid arthritis and one with diffuse alopecia. The table includes the library clone identifier, gene name, database accession number, and the relative intensity units measured for each of the duplicate clones on the chip.

No.	Clone MPMGp800	Gene name	Accession Number	Arthritis serum43	Arthritis serum44	Diffuse alopecia
1	A21618	Autoantigen p69	L21181	37.347	172.012	149.821
				48.309	180.647	158.224
2	G02617	Tumor suppressor p33 ING homolog	AF044076	38.771	60.704	190.685
				43.064	91.959	194.042
3	I17618	cDNA FLJ20427, unknown function	AK000434	23.381	103.153	142.685
				27.347	128.798	149.877
4	O03614	Unknown		108.744	57.624	34.262
				126.624	60.923	26.532
5	B15613	Homolog to mouse synaptotagmin VI	NM018800	14.178	123.594	32.764
				22.172	146.278	36.381
6	F11617	Diglycerol kinase ζ	XM012056	142.813	61.501	29.240
				147.589	85.869	27.069
7	K20617	Ubiquitin B (control clone)	NM018955	35.369	52.203	25.680
				40.484	72.644	28.148
8	L16612	RAS association domain family 1	XM003273.2	79.750	135.898	39.501
				88.095	147.211	50.212

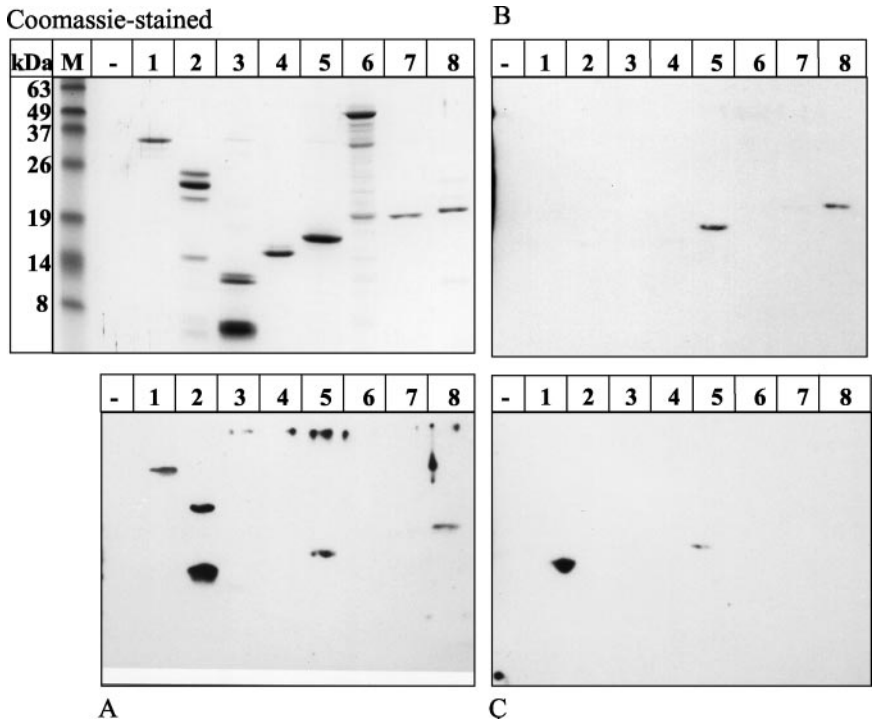

FIG. 5. **Western immunoblotting of purified candidate autoantigens.** *Left side top*, purified candidate autoantigens on Coomassie-stained SDS-PAGE. *Lanes 1–8*, p69; tumor suppressor p33 ING homologue, cDNA FLJ20427, unknown, mouse synaptotagmin VI homologue, diglycerol kinase ζ, ubiquitin B (control clone), and RAS assocation domain family. *A*, Western blot of candidate autoantigens probed with serum from patient with diffuse alopecia. *B*, Western blot of candidate autoantigens probed with serum from patient with alopecia areata. *C*, Western blot of candidate autoantigens probed with the secondary antibody (control).

the background threshold shows that most of them encode their gene product in the wrong reading frame with respect to the $RGSH_6$ epitope. Only a minority of clones exhibiting a low relative intensity expressed its cDNA insert in the correct reading frame. Sequence analysis revealed that these clones encode for either highly structured proteins or large proteins that may be difficult to express by *E. coli* (32). Clones showing a higher relative intensity of the expressed proteins are mainly in the correct reading frame, but some wrong frame clones resulting in small 5–8-kDa peptides have also been observed, indicating that some of the small peptides are not fully degraded by *E. coli*. The use of purified proteins allows certain discrimination between correct and wrong frame clones.

Sequence analysis of 2303 of the 2413 clones was performed, and the reading frame of 1602 was determined and gave the following results: 880 (55%) clones have their cDNA insert in the $RGSH_6$-tag corresponding reading frame, including 382 (43%) clones comprising full-length proteins, and 724

(45%) clones that show a different reading frame to the frame determined by the $RGSH_6$-tag. Using stringent sequence analysis criteria, no reading frame was determined for 701 clones.

We have demonstrated the characterization of antibody specificity on high-density protein microarrays, as well as their potential application in profiling the antibodies repertoire in sera from patients with autoimmune diseases. In order to minimize all surface-related effects that might cause nonspecific binding or unspecific folding of the deposited proteins on conventional slides, we worked with a highly aqueous gel. This also allows the addition of reagents that may increase binding or change the surrounding environment to a more physiological state. A related three-dimensional surface was developed by Mirzabekov and coworkers, who have used a gel photo or persulfate-induced copolymerization technique to produce functional protein microarrays on polyacrylamide gel pads (33). Furthermore, an additional advantage of the three-dimensional structure of the gel matrix, when compared with conventional two-dimensional slides, is the higher protein binding capacity and their high detection sensitivity (34). Other methods to generate protein microarrays involve protein immobilization on a flat, two-dimensional surface by either covalent coupling to a cross-linker attached to the surface (35–37) or noncovalent interaction to an immobilized biomolecule such as biotin (38). Recently, oriented immobilization of proteins either on chlorinated glass slides (39) via affinity tags (40) or by biotinylation of capture molecules and their immobilization on streptavidin coated supports (41) have been described.

For the generation of antibody arrays, which consist of immobilized antibodies used to detect different proteins, for example in body fluids or tissue extracts, it is increasingly important to analyze these antibodies for their specificity against a large number of different proteins. We have analyzed and compared the specificity of two mouse monoclonal antibodies, anti-GAPDH and anti-HSP90β, respectively, using our protein microarrays. It has been show that both antibodies recognize their antigen with high specificity, but additional proteins were also detected, indicating a certain cross-reactivity. Both antibodies were designed for their use in Western immunoblotting and, therefore, recognize linear epitopes. However, because refolding of immobilized, denatured proteins was shown to occur in a single step using Tris-buffered saline (42), which is also used here in the binding assay, a partial refolding of the immobilized proteins during incubation cannot be excluded. Therefore, the possibility remains that the resulting structural epitopes might mimic the antigenic conformation. In respect to antibodies immobilized on antibody arrays, best results will be obtained using antibodies, which were previously characterized. In addition, these results show that this Uniclone set is indeed nonredundant, because the antibodies characterized here recognize housekeeping genes, which were highly redundant in the original cDNA library (14).

In the proof of principle experiment to profile the antibody repertoire in serum, we chose sera from two different patients with rheumatoid arthritis and one serum from a patient showing characteristic symptoms of alopecia. Protein p69 was recognized by the sera from patients with different autoimmune diseases (arthritis and alopecia). Interestingly, p69 is also known to be a candidate autoimmune target in type 1 diabetes as the recombinant protein was recognized by autoantibodies and T cells from diabetic children (43). These results suggest that this protein may be involved in a more general way in autoimmune diseases. The microarray data was confirmed by Western immunoblot analysis of equal amounts of the purified proteins (Fig. 5) with the diffuse alopecia patient sera. This serum detected specifically the autoantigen p69 and the tumor suppressor p33 ING homologous protein, whereas the gene product of cDNA FLJ20427 was not detected on the Western immunoblot (Fig. 5*A*). This might be due to changes in the conformation after SDS treatment or to effects of the immobilization either on the chip or during Western immunoblot procedure. Additionally, weaker signals were obtained for the homologue of mouse synaptotagmin VI and the protein belonging to the RAS association domain family I. Both of these proteins were also recognized by the serum of the patient with characteristics specific for alopecia areata, whereas autoantigen p69 and tumor suppressor p33 ING homologous protein were not detected, consistent with the clinical diagnosis of different types of hair loss (Fig. 5*B*). This suggests that protein arrays containing a specific subset of proteins may be used for subtyping of diseases. A more in-depth analysis of the selected autoimmune diseases, including the characterization of sera from large patient and control cohorts selected after clinical diagnosis, is currently under way.

In summary, using our protein microarrays consisting of a large number of human proteins, the specificity and cross-reactivity of antibodies can be characterized. The detected proteins can easily be identified by mass spectrometry or DNA sequencing, as they are derived from clones of an arrayed cDNA expression library. This approach can also be applied to determine the binding specificity of antibodies, which were previously unknown, or to profile the binding of mixtures of antibodies, such as are found in sera of patients with autoimmune diseases.

Acknowledgments—We thank Prof. Dr. Ulrike Blume-Peytavi and Dr. Dorothee Hagemann for providing the patient sera.

* This work was funded by Bundesministerium für Bildung und Forschung (BioFuture 0311018 and DHGPII 0311323) and the Max-Planck-Gesellschaft. D. J. C acknowledges funding from the Health Education Authority and Science Foundation Ireland, Dublin, Ireland. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

** To whom correspondence should be addressed: Centre for Human Proteomics, Royal College of Surgeons in Ireland, 123 St. Ste-

phen's Green, Dublin 2, Ireland. Tel.: 353-(1)-402-8550; Fax: 353-(1)-402-2453; E-mail: chp@rcsi.ie.

REFERENCES

1. Epstein, C. B., and Butow, R. A. (2000) Microarray technology—Enhanced versatility, persistent challenge. *Curr. Opin. Biotechnol.* **11,** 36–41
2. Lockhart, D. J., and Winzeler, E. A. (2000) Genomics, gene expression and DNA arrays. *Nature* **405,** 827–836
3. Blohm, D. H., and Guiseppi-Elie, A. (2001) New developments in microarray technology. *Curr. Opin. Biotechnol.* **12,** 41–47
4. Meier-Ewert, S., Maier, E., Ahmadi, A., Curtis, J., and Lehrach, H. (1993) An automated approach to generating expressed sequence catalogues. *Nature* **361,** 375–376
5. Lennon, G. G., and Lehrach, H. (1991) Hybridization analysis of arrayed cDNA libraries. *Trends Genetics* **7,** 314–317
6. Schena, M., Heller, R. A., Theriault, T. P., Konrad, K., Lachenmeier, E., and Davis, R. W. (1998) Microarrays: Biotechnology's discovery platform for functional genomics. *Trends Biotechnol.* **16,** 301–306
7. Eickhoff, H., Schuchhardt, J., Ivanov, I., Meier-Ewert, S., O'Brien, J., Malik, A., Tandon, N., Wolski, E. W., Rohlfs, E., Nyarsik, L., Reinhardt, R., Nietfeld, W., and Lehrach, H. (2000) Tissue gene expression profiling using arrayed normalized cDNA libraries. *Genome Res.* **10,** 1230–1240
8. Walter, G., Büssow, K., Cahill, D., Lueking, A., and Lehrach, H. (2000) Protein arrays for gene expression and molecular interaction screening. *Curr. Opin. Microbiol.* **3,** 298–302
9. Cahill, D. et al (2001) in *Proteomics* (Pennington, S. and Dunn, M., eds.) pp. 1–20, BIOS Scientific Publishers, Oxford
10. Pohlner, K., Portig, I., Pankuweit, S., Lottspeich, F., and Maisch, B. (1997) Identification of mitochondrial antigens recognized by antibodies in sera of patients with idiopathic dilated cardiomyopathy by two-dimensional gel electrophoresis and protein sequencing. *Am. J. Cardiol.* **80,** 1040–1045
11. Latif, N., Baker, C. S., Dunn, M. J., Rose, M. L., Brady, P., and Yacoub, M. H. (1993) Frequency and specificity of antiheart antibodies in patients with dilated cardiomyopathy detected using SDS-PAGE and western blotting. *J. Am. Coll. Cardiol.* **22,** 1378–1384
12. Greiner, J., Ringhoffer, M., Simikopinko, O., Szmaragowska, A., Huebsch, S., Maurer, U., Bergmann, L., and Schmitt, M. (2000) Simultaneous expression of different immunogenic antigens in acute myeloid leukemia. *Exp. Hematol.* **28,** 1413–1422
13. Cahill, D. (2000) Protein arrays: A high-throughput solution for proteomics research? in *Proteomics: Current Trends Supplement* (Blackstock, W., and Mann, M., eds) pp. 47–51, Elsevier Science, Ltd., London
14. Büssow, K., Cahill, D., Nietfeld, W., Bancroft, D., Scherzinger, E., Lehrach, H., and Walter, G. (1998) A method for global protein expression and antibody screening on high-density filters of an arrayed cDNA library. *Nucleic Acids Res.* **26,** 5007–5008
15. Lehrach, H., Bancroft, D., and Maier, E. (1997) Robotics, computing, and biology. *Interdisciplinary Sci. Rev.* **22,** 37–44
16. Lueking, A., Horn, M., Eickhoff, H., Bussow, K., Lehrach, H., and Walter, G. (1999) Protein microarrays for gene expression and antibody screening. *Anal. Biochem.* **270,** 103–111
17. Büssow, K., Nordhoff, E., Lübbert, C., Lehrach, H., and Walter, G. (2000) A human cDNA library for high-throughput protein expression screening. *Genomics* **65,** 1–8
18. Schmidt, F., Lueking, A., Nordhoff, E., Gobom, J., Klose, J., Seitz, H., Egelhofer, V., Eickhoff, H., Lehrach, H., and Cahill, D. J. (2002) Generation of minimal protein identifiers of proteins from two-dimensional gels and recombinant proteins. *Electrophoresis* **23,** 621–625
19. Herwig, R., Schmitt, A. O., Steinfath, M., O'Brien, J., Seidel, H., Meier-Ewert, S., Lehrach, H., and Radelof, U. (2000) Information theoretical probe selection for hybridisation experiments. *Bioinformatics* **16,** 890–898
20. Yang, Y. H., Dudoit, S., Luu, P., and Speed, T. P. (2000) *Technical Report*, University of California, Berkley, CA
21. Schuchhardt, J., Beule, D., Malik, A., Wolski, E., Eickhoff, H., Lehrach, H., and Herzel, H. (2000) Normalization strategies for cDNA microarrays. *Nucleic Acids Res.* **28,** E47
22. Hochuli, E., Dobeli, H., and Schacher, A. (1987) New metal chelate adsorbent selective for proteins and peptides containing neighboring histidine residues. *J. Chromatogr.* **411,** 177–184
23. Borrebaeck, C. A. (2000) Antibodies in diagnostics—From immunoassays to protein chips. *Immunol. Today* **21,** 379–382
24. Robinson, W. H., DiGennaro, C., Hueber, W., Haab, B. B., Kamachi, M., Dean, E. J., Fournel, S., Fong, D., Genovese, M. C., de Vegvar, H. E., Skriner, K., Hirschberg, D. L., Morris, R. I., Muller, S., Pruijn, G. J., van Venrooij, W. J., Smolen, J. S., Brown, P. O., Steinman, L., and Utz, P. J. (2002) Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nat. Med.* **8,** 295–301
25. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402
26. MacBeath, G., and Schreiber, S. L. (2000) Printing proteins as microarrays for high-throughput function determination. *Science* **289,** 1760–1763
27. Haab, B. B., Dunham, M. J., and Brown, P. O. (2001) Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biology* **2,** research0004.0001–0004.0013
28. Holz, C., Lueking, A., Bovekamp, L., Gutjahr, C., Bolotina, N., Lehrach, H., and Cahill, D. J. (2001) A human cDNA expression library in yeast enriched for open reading frames. *Genome Res.* **11,** 1730–1735
29. Lueking, A., Holz, C., Gotthold, C., Lehrach, H., and Cahill, D. (2000) A system for dual protein expression in *Pichia pastoris* and *Escherichia coli*. *Protein Expr. Purif.* **20,** 372–378
30. Albala, J. S., Franke, K., McConnell, I. R., Pak, K. L., Folta, P. A., Rubinfeld, B., Davies, A. H., Lennon, G. G., and Clark, R. (2000) From genes to proteins: High-throughput expression and purification of the. *J. Cell. Biochem.* **80,** 187–191
31. Lueking, A., Horn, S., Lehrach, H., and Cahill, D. J. (2003) A dual-expression vector allowing expression in *E. coli* and *P. pastoris*, including new modifications. *Methods Mol. Biol.* **205,** 31–42
32. Makrides, S. C. (1996) Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol. Rev.* **60,** 512–538
33. Arenkov, P., Kukhtin, A., Gemmell, A., Voloshchuk, S., Chupeeva, V., and Mirzabekov, A. (2000) Protein microchips: Use for immunoassay and enzymatic reactions. *Anal. Biochem.* **278,** 123–131
34. Angenendt, P., Glökler, J., Murphy, D., Lehrach, H., and Cahill, D. J. (2002) Towards antibody microarrays: A comparison of current microarray support materials. *Anal. Biochem.* **309,** 253–260
35. Jones, V. W., Kenseth, J. R., Porter, M. D., Mosher, C. L., and Henderson, E. (1998) Microminiaturized immunoassays using atomic force microscopy and compositionally patterned antigen arrays. *Anal. Chem.* **70,** 1233–1241
36. Mendoza, L. G., McQuary, P., Mongan, A., Gangadharan, R., Brignac, S., and Eggers, M. (1999) High-throughput microarray-based enzyme-linked immunosorbent assay (ELISA). *BioTechniques* **27,** 778–780, 782–776, 788
37. Mooney, J. F., Hunt, A. J., McIntosh, J. R., Liberko, C. A., Walba, D. M., Rogers, C. T. (1996) Patterning of functional antibodies and other proteins by photolithography of silane monolayers. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 12287–12291
38. Yang, Z., Frey, W., Oliver, T., and Chilkoti, A. (2000) Light-activated affinity micropatterning of proteins on self-assembled monolayers on gold. *Langmuir* **16,** 1751–1758
39. Hergenrother, P. J., Depew, K. M., and Schreiber, S. L. (2000) Small-molecule microarrays: Covalent attachment and screening of alcohol-containing small molecules on glass slides. *J. Am. Chem. Soc.* **122,** 7849–7850
40. Zhu, H., and Snyder, M. (2001) Protein arrays and microarrays. *Curr. Opin. Chem. Biol.* **5,** 81–91
41. Peluso, P., Wilson, D. S., Do, D., Tran, H., Venkatasubbaiah, M., Quincy, D., Heidecker, B., Poindexter, K., Tolani, N., Phelan, M., Witte, K., Jung, L. S., Wagner, P., and Nock, S. (2003) Optimizing antibody immobilization strategies for the construction of protein microarrays. *Anal. Biochem.* **312,** 113–124
42. Holzinger, A., Phillips, K. S., and Weaver, T. E. Single-step purification/solubilization of recombinant proteins: Application to surfactant protein B. *BioTechniques* ***20,*** 804–806, 808
43. Miyazaki, I., Gaedigk, R., Hui, M. F., Cheung, R. K., Morkowski, J., Rajotte, R. V., and Dosch, H. M. (1994) Cloning of human and rat p69 cDNA, a candidate autoimmune target in type 1 diabetes. *Biochim. Biophys. Acta* **1227,** 101–104