Purification and Characterization of the Mammaglobin/Lipophilin B Complex, a Promising Diagnostic Marker for Breast Cancer†

Darrick Carter,*,‡ John F. Douglass,‡ Charisa D. Cornellison,‡ Marc W. Retter,‡ Jeffrey C. Johnson,‡ Angela A. Bennington,‡ Timothy P. Fleming,§ Steven G. Reed,‡ Raymond L. Houghton,‡ Deborah L. Diamond,|| and Thomas S. Vedvick‡

Department of Antigen Discovery, Corixa Corp., 1124 Columbia Street, Suite 200, Seattle, Washington 98104, Department of Surgery, Washington University School of Medicine, St. Louis, Missouri 63130, and Ciphergen Biosystems Inc., Fremont, California 94555

Received November 28, 2001; Revised Manuscript Received March 29, 2002

ABSTRACT: Mammaglobin, a promising diagnostic marker for breast cancer, forms a covalent complex with lipophilin B. mRNA levels for each component of the complex were determined for a number of breast tumors and normal tissues, and correlation of message expression was highly significant between mammaglobin and lipophilin B ($p < 0.0001$). The complex was purified by both standard biochemical techniques and immunoaffinity chromatography. N-Terminal sequencing revealed that mammaglobin and lipophilin B are processed as predicted by cleavage of their signal sequence after amino acids 19 and 21, respectively. Three molecular masses—representing the fully glycosylated form, the complex without one of the carbohydrate chains, and the deglycosylated proteins—are detected by ProteinChip array SELDI-TOF mass spectrometry after partial enzymatic deglycosylation. This is consistent with the two predicted N-linked glycosylation sites in the primary sequence of mammaglobin and each site having an attached sugar of ~3500 Da. Reducing agents release lipophilin B from mammaglobin, and the free peptides are seen at their predicted molecular masses in the deglycosylated complex. Molecular modeling, secondary structure prediction, and circular dichroism indicate that the complex is a small α-helical globule that has three disulfide bridges and a carbohydrate chain at each pole. LC-ESI-MS shows that mammaglobin and lipophilin B are bonded in a head to tail orientation. This work describes the biochemistry of the mammaglobin/lipophilin B complex and lays the framework for use of this complex as a novel protein-based serological marker for breast cancer.

The use of markers to detect tumors is becoming increasingly more important for the early detection of small and spreading cancers with the advent of routine ELISA[1] and PCR procedures. These methods have been very successful in diseases for which specific markers, such as the prostate-specific antigen for prostate cancer, are known. Mammaglobin has been shown to be overexpressed in breast cancers (*1*) and is propounded as a highly specific marker for breast cancer (*2*). In recent studies, 81% of breast cancers were immunopositive for mammaglobin protein expression, and 23% of tumors showed a 10-fold increase in mRNA levels above those found in normal breast cells (*1*, *3*). Overall, >80% of breast tumors are positive for mammaglobin mRNA and can be detected using real-time PCR (*4*). Despite the impressive amount of work on gene expression of mammaglobin, very little is known about the native protein. Only recently has an association of mammaglobin with lipophilin B been reported, which was based on coelution profiles on dot blots (*5*, *6*). Preliminary characterization in these papers indicates there may be higher order quaternary structures formed by the covalent duplex, although extensive studies have been hampered by the lack of highly purified material.

Mammaglobin is expressed as a small polypeptide of 93 amino acids in length with a predicted molecular mass of 10,500 kDa and a putative cleavage site at amino acid 19 (*1*). It is homologous to lipophilin C (also known as mammaglobin B) with 52% identity on the amino acid level (*7*) and is similar to the rat prostatic binding protein component C3 and rabbit uteroglobin (*8*, *9*). These proteins are known to exist as covalent homo- or heterodimers, which are linked by disulfides (*10*, *11*), can form higher order multimers, and can bind to aromatic compounds such as polychlorinated biphenyls and progesterone (*12*, *13*). Functions attributed to the uteroglobin family are manyfold including immunosuppression by inhibiting proliferation of

† This work was funded in part by National Cancer Institute Grants CA-75794, CA-86673, and CA-76227.

* Corresponding author [telephone (206) 754-5957; fax (206) 754-5715; e-mail carter@corixa.com].

‡ Corixa Corp.

§ Washington University School of Medicine.

|| Ciphergen Biosystems Inc.

[1] Abbreviations: ELISA, enzyme-linked immunosorbent assay; ESI, electrospray ionization; DTT, dithiothreitol; DMEM, Dulbecco's Modified Eagle Medium; SELDI, surface-enhanced laser desorption ionization; TOF, time of flight; MALDI, matrix-assisted laser desorption ionization; CHCA, α-cyano-4-hydroxycinnamic acid; LC, liquid chromatography; PBS, phosphate-buffered saline; ELISA, enzyme-linked immunosorbant assay; MS, mass spectrometry; PAGE, polyacrylamide gel electrophoresis; SDS, sodium dodecyl sulfate; Tris, tris[hydroxymethyl]aminomethane; PCR, Polymerase Chain Reaction; CD, circular dichroism.

10.1021/bi0159884 CCC: $22.00
Published on Web 05/02/2002

lymphocytes and decreasing IL-2 production (*14*), induction of autoimmunity while functioning as both a humoral and cellular antigen (*15*), anti-inflammatory action of peptides derived from uteroglobin proteins (*16*), cytokines in an as-yet undefined receptor-mediated pathway (*17*), and inhibition of phospholipase A2 (*18*), although the specificity of this inhibition has been debated (*19*). A straightforward picture has not emerged yet of the function of these complexes, but the consensus appears to be that they bind aromatic molecules such as steroids and biphenyls when reduced and are involved in regulating the immune system (*12*, *13*, *20*). In this paper we report the purification of the native mammaglobin/lipophilin B complex from a breast cancer tumor cell line supernatant and the biochemical characterization of this complex. These findings will facilitate the study of the role of this complex in healthy and cancerous states and the development of diagnostic assays for breast cancer.

EXPERIMENTAL PROCEDURES

Chemicals and Reagents. Tris[hydroxymethyl]aminomethane (Tris), sodium chloride, ammonium acetate, phosphate-buffered saline (PBS), glycine, methanol, silver nitrate, and sodium thiosulfate were obtained from Sigma (St. Louis, MO). Acetonitrile was from Mallinckrodt Laboratory Chemicals (Phillipsburg, NJ). All other reagents were of the highest quality commercially available. All tissue samples were obtained after Institutional Review Board (IRB) approval at each site. Corixa Corp. also has IRB approval for antigen discovery research.

Quantitative Real-Time PCR. The specificity and sensitivity of the different genes was determined using quantitative PCR analysis. Breast metastases, primary breast tumors, benign breast disorders, and normal breast tissue along with other normal tissues and tumors were tested in quantitative (real-time) PCR. This was performed either on the ABI 7700 Prism or on a GeneAmp 5700 sequence detection system (PE Biosystems, Foster City, CA). The ABI 7700 system uses a forward and a reverse primer in combination with a specific probe with a 5′ fluorescent reporter dye at one end and a 3′ quencher dye at the other end (Taqman). During PCR, using the Taq DNA polymerase with 5′−3′ nuclease activity, the probe is cleaved and begins to fluoresce, allowing the reaction to be monitored by the increase in fluorescence. The GeneAmp 5700 system uses a fluorescent dye, SYBR green, that binds only to double-stranded DNA and the same forward and reverse primers as the ABI 7700 instrument. Matching primers and fluorescent probes were designed for each of the genes according to the primer express program (PE Biosystems). The primers used for mammaglobin detection were Mamm 2f, 5′-TGCCATAGATGAATTGAAGGAATG-3′, and Mamm 2r, 5′-TGTCATATATTAATTGCATAAACACCTCA-3′; for lipophilin B the primers were LipoBf, 5′-TGCCCCTCCGGAAGCT-3′, and LipoBr, 5′-CGTTTCTGAAGGGACATCTGATC-3′. Primers and probes so produced were used in the universal thermal cycling program in real-time PCR. They were titrated to determine the optimal concentrations using a checkerboard approach. A pool of cDNA from target tumors was used in this optimization process. The reaction was performed in 25 μL volumes. The final probe concentration in all cases was 160 nM. dATP, dCTP, and dGTP were at 0.2 mM, and dUTP was at 0.4 mM. Amplitaq gold and Amperase UNG (PE Biosystems) were used at 0.625 and 0.25 unit per reaction. $MgCl_2$ was at a final concentration of 5 mM. Trace amounts of glycerol, gelatin, and Tween 20 (Sigma Chemical Co.) were added to stabilize the reaction. Each reaction contained 2 μL of diluted template. The cDNA from RT reactions prepared as above is diluted 1:10 for genes of interest and 1:100 for β-actin. Primers and probes for β-actin (PE Biosystems) were used in a similar manner to quantitate the presence of β-actin in the samples. In the case of the SYBR green assay the reaction mix (25 μL) included 2.5 μL of SYBR green buffer, 2 μL of cDNA template, and 2.5 μL each of the forward and reverse primers for the gene of interest. This mix also contained 3 mM $MgCl_2$, 0.25 unit of AmpErase UNG, 0.625 unit of Amplitaq gold, 0.08% glycerol, 0.05% gelatin, 0.0001% Tween 20, and 1 mM dNTP mix. In both formats, 40 cycles of amplification were performed. To quantify the amount of specific cDNA (and hence initial mRNA) in the sample, a standard curve is generated for each run using a plasmid containing the gene of interest. Standard curves were generated using the Ct values determined in the RT-PCR that related to the initial cDNA concentration used in the assay. Standards ranging from 20 to 2×10^6 copies of the gene of interest were used for this purpose. In addition, a standard curve is generated for the housekeeping gene β-actin ranging from 200 fg to 2000 pg to enable normalization to a constant amount of β-actin. This allows the evaluation of the overexpression levels seen with each of the genes (*21*, *22*).

Cell Culture. The adherent cell line MDA-MB-415 was used as it demonstrates high levels of secreted mammaglobin protein when grown under the conditions described below. It was originally isolated from a female patient diagnosed with breast adenocarcinoma and is available through the American Type Culture Collection (ATCC No. HTB-128). Cells were initially grown in DMEM (Life Technologies Inc., Rockville, MD) supplemented with 10% fetal bovine serum (Hyclone, Logan, UT). Once the cells were confluent, the serum-containing medium was removed; the cells were rinsed with PBS, and low-serum medium was added. The low-serum medium consisted of a mix of 50% SFX-CHO (Hyclone) supplemented with 2 mM GlutaMAX (Life Technologies Inc.) and 50% Opti-MEM (Life Technologies Inc.). Insulin−transferrin−selenium A (Life Technologies Inc.) was added for a final concentration of 1×. The cells were grown in this medium for a period of several weeks, with weekly applications of fresh medium. Supernatants were collected and pooled for subsequent mammaglobin isolation.

Protein Purification. Mammaglobin complex was purified by two different methods: Initially, 3 L of serum-free culture supernatant was concentrated by ultrafiltration through an XK10 membrane (Amicon) to ~300 mL. These were dialyzed against numerous changes of 10 mM Tris (pH 8) at 4 °C over 2 days. The dialysate was loaded with an ÄKTA explorer 100 (Amersham Pharmacia Biotech AB, Uppsala, Sweden) onto a 50 mL Macro-Prep High Q (Bio-Rad, Hercules, CA) anion exchange column and eluted with a complex gradient from 100% buffer A (10 mM Tris, pH 8.0) to 20% buffer B (10 mM Tris, pH 8.0, 1 M NaCl) followed by a flattened incline to 55% buffer B followed by a final step to 100% buffer B. Fractions were run on a 4−20% SDS−PAGE gradient gel and analyzed by Western blotting. Positive fractions were pooled and dialyzed against water

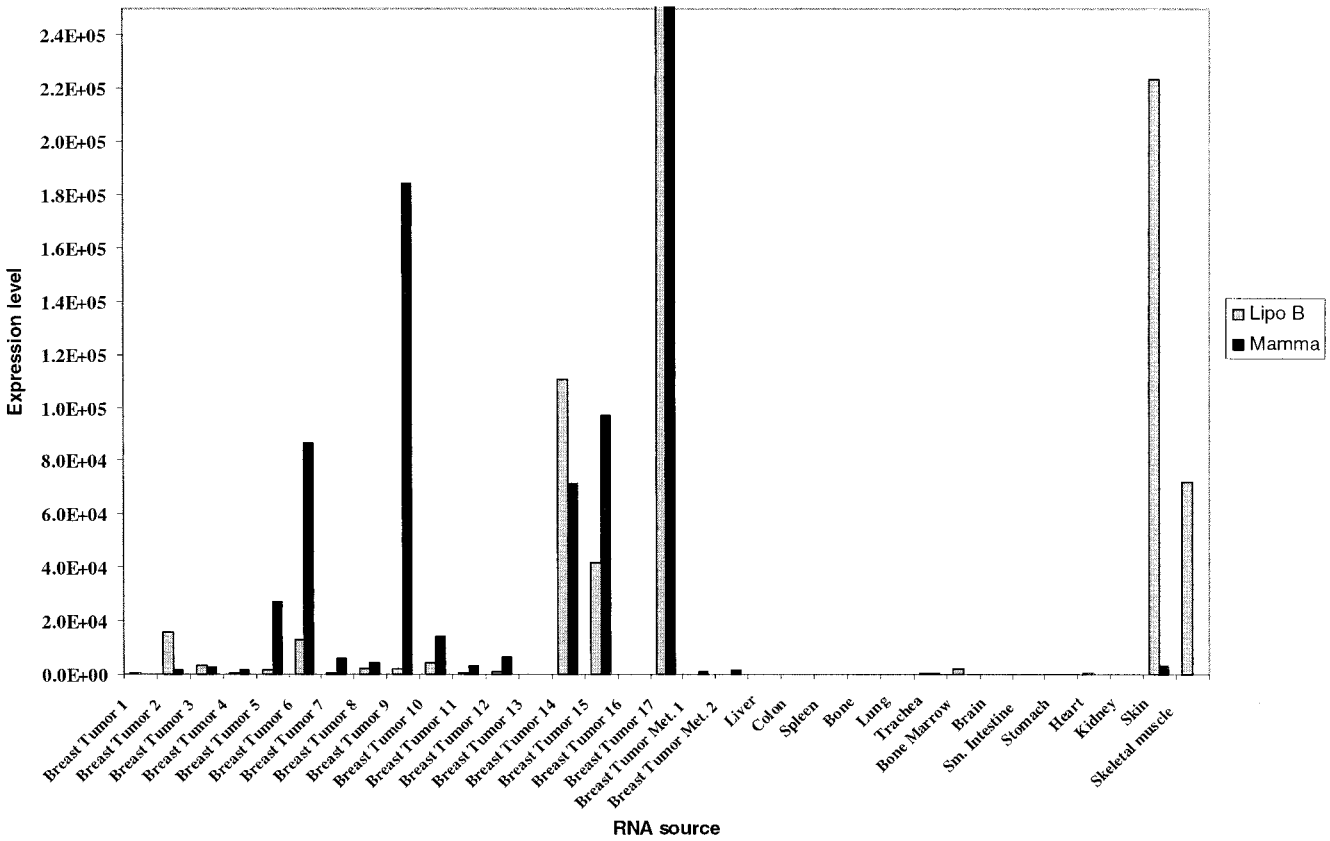


FIGURE 1: Transcription of mammaglobin mRNA correlated with transcription of lipophilin B. The number of transcripts as determined by RT-PCR is shown for a variety of breast tumors and tissue types. There is a significant correlation of message expression between mammaglobin and lipophilin B in cancers, whereas some normal tissues such as skeletal muscle express lipophilin B but not mammaglobin.

overnight. Desalted, pooled material was loaded with an ÄKTA Explorer 100 onto a reversed phase column, Resource 15 RPC (Amersham Pharmacia Biotech AB), and protein was eluted using a complex gradient of 10 mM ammonium acetate in water (pH 7.0) against 100% acetonitrile. The gradient ran from 0 to 20% buffer B, was held at 20%, was then run slowly to 45% buffer B, held there, then sharply increased to 65% buffer B, and stepped to 100% buffer B. Again, fractions were analyzed by Western blotting, and positive fractions were pooled and lyophilized. This protein was used to generate a monoclonal antibody, RO28, in rabbits using the selected lymphocyte antibody method (*23*). In this procedure single antibody-specific B cells are isolated using antigen-specific hemolytic plaque assays. Heavy and light chain cDNAs are prepared from such cells using RT-PCR and cloned into myeloma or CHO cells to produce antibody. This procedure enables the generation of high-affinity and very specific antibodies. RO28 specifically recognizes the native complex, although it does bind to synthetic peptides from lipophilin B, albeit with much lower affinity. One milligram of RO28 was immobilized on an Aminolink resin (Pierce, Rockford, IL) according to the protocol provided. Again, culture supernatant was chromatographed over this resin, washed with PBS, and eluted in 0.2 M glycine (pH 3.5). Fractions were immediately neutralized with 1 M Tris. The protein was then chromatographed by a reversed phase chromatography similar to the one described above. Purified protein was lyophilized, dissolved in PBS, and stored frozen.

Deglycosylation of Purified Complex. The purified complex was lyophilized for 2 days to complete dryness and then chemically deglycosylated with trifluoromethanesulfonic acid using the GlycoFree deglycosylation kit (Glyko Inc., Novato, CA). Additionally, mammaglobin was enzymatically deglycosylated with PNGase F (Sigma) in the absence of reducing agents.

SDS−PAGE; PVDF- and Western Blotting. SDS−PAGE was performed according to the method of Laemmli (*24*). Samples were diluted 1:1 in Laemmli sample buffer (Invitrogen, Carlsbad, CA) and boiled for 5 min. Proteins were then loaded on 4−20% acrylamide gradient gels (Bio-Rad) and run at a constant 250 V for 30 min. Gels were blotted by transferring to a Sequi-Blot PVDF membrane (Bio-Rad) in 50 mM Tris base, 40 mM glycine, and 20% methanol using a Bio-Rad Trans-Blot cell. Blots were transferred for 100 min at 100 V. For Western blots, PVDF membranes were then blocked for 1 h in PBS containing 0.5 M NaCl and 1% Tween-20 (Sigma). Next, the membranes were washed three times for 10 min with the same buffer containing 0.1% Tween-20 (PBSST). Membranes were probed with 1 μg/mL of 2D3, a monoclonal antibody to recombinant mammaglobin (Corixa Corp., Seattle, WA), overnight and washed again as above. A horseradish peroxidase conjugated goat anti-rabbit IgG was used as a secondary antibody, developed with ECL solution (Amersham Pharmacia Biotech AB), and visualized using scientific imaging film (Kodak, Rochester, NY).

Protein Detection. Gels or blots were stained with either silver (*25*), Coomassie Brilliant Blue R250, or the Glyco-Pro glycoprotein detection kit (Sigma) as described (*26*).

N-Terminal Sequencing. Amino-terminal sequence data were obtained from purified proteins that were dried directly

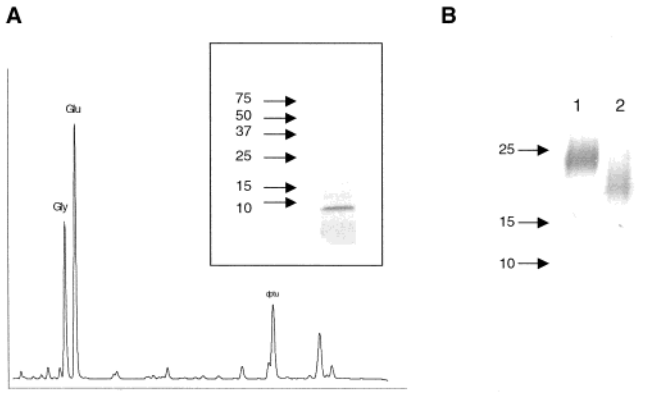


FIGURE 2: Purity and carbohydrate staining of the complex: (A) First residue trace from N-terminal sequencing of purified complex. Protein samples after purification were analyzed by N-terminal sequence analysis on a Procise 494 N-terminal sequencing apparatus. The purified complex contained <10% other proteins, notably serum albumin. The molar ratios of mammaglobin to lipophilin B were calculated to be 1.18 ± 0.20 on the basis of four sequencing runs, consistent with a 1:1 molar ratio of mammaglobin to lipophilin B in the complex. The larger apparent size of the lipophilin B peak is largely due to the more efficient conversion of glutamyl residues when compared to glycinyl residues during the N-terminal sequencing process. Further sequencing reveals the signal sequences are removed at the predicted sites: mammaglobin begins at amino acid 20 (Gly-Ser-Gly...), and lipophilin B begins at amino acid 22 (Glu-Phe-Pro...). (Inset) Coomassie staining of reduced, purified complex. Only lipophilin B is visible in this gel as mammaglobin is highly glycosylated and does not stain well. (B) Mammaglobin complex glycosylated and associated by disulfide bonds. An SDS−PAGE gel of complex was run reduced and nonreduced and stained with the Glyco-Pro glycoprotein detection kit. A parallel gel was run and blotted onto a PVDF membrane, stained similarly, and bands were N-terminally sequenced: (lane 1) native complex stains well with the carbohydrate-specific stain (band contained both mammaglobin and lipophilin B sequence in a 1:1 molar ratio); (lane 2) reduced complex also stains well with the carbohydrate-specific stain, but no lipophilin B signal is detected in this band.

onto TFA-treated glass fiber filters (Perkin-Elmer/Applied Biosystems Division) or from samples which were separated on SDS−PAGE and electroblotted onto Sequi-Blot PVDF membranes according to the method of Matsudaira (*27*). The membranes were stained with the Glyco-Pro glycoprotein detection kit (Sigma) or Coomassie Brilliant Blue to visualize the protein bands that were excised with a clean razor. Traditional Edman degradation sequence analysis was performed using a Perkin-Elmer/Applied Biosystems Division Procise 494 protein sequencer.

Molecular Modeling and Secondary Structure Prediction. The Swiss-Model package (*28*) of modeling software was used to generate a molecular model of the mammaglobin/lipophilin B complex. The coordinates of a related protein from the uteroglobin family, rat uteroglobin (*29*) (PDB ID: 1utr), were used as a template. The amino acid sequences of mammaglobin and lipophilin B each were aligned with chains from the coordinate file using the method of Hein (*30*) in the MegAlign program provided with the DNASTAR sequence analysis software package (DNASTAR Inc., Madison, WI). Each alignment was manually transferred to the Swiss Pdb Viewer provided with the Swiss-Model package and sent for automated modeling. The two coordinate files received were then combined to display the model of the full complex. Secondary structure was then extracted using the program Stride (*31*) and compared with the predicted secondary structure output from nnpredict (*32*, *33*) run in the all-alpha mode.

Circular Dichroism (CD). Mammaglobin complex was dialyzed against water and the concentration of the resulting material determined. A CD spectrum was then collected on a JASCO J-810 CD spectropolarimeter using a 0.1 cm path length cuvette at room temperature.

Liquid Chromatography−Electronspray Ionization Mass Spectroscopy (LC−ESI MS). Samples were prepared by adding urea to 50 μg of purified mammaglobin to a final concentration of 8 M and incubating at 37 °C for 30 min. Mammaglobin was then dialyzed against 100 mM ammonium bicarbonate using Slide-A-Lyzer mini dialysis units with a 3500 Da molecular mass cutoff (Pierce). TPCK-trypsin was added at 1:50 w/w and incubated at room temperature overnight. The sample was lyophilized to a final volume of 100 μL on a Savant Speed-Vac (Savant Instruments Inc., Holbrook, NY). Reduction of sulfhydryls was accomplished by adding DTT to a final concentration of 10 mM and incubating at 37 °C for 1 h. Tryptic peptides were run on a model 140C microgradient HPLC (Applied Biosystems, Foster City, CA). Five micrograms of sample was loaded onto a 1.0 mm Vydac C18 column (Vydac, Hesperia, CA) and eluted with a linear gradient from 0 to 100% acetonitrile with 1% acetic acid at 50 μL/min. Peptides eluted from the reversed phase column were analyzed directly on an ion trap LCQ mass spectrometer (Finnigan, San Jose, CA). Mass spectra were acquired over a mass range of 150−2000 Da in MS mode.

Surface-Enhanced Laser Desorption Ionization (SELDI) Time of Flight (TOF) Mass Spectrometry. Native and enzymatically deglycosylated mammaglobin complexes were analyzed by SELDI-TOF mass spectrometry in a Ciphergen Protein Biology System II (PBS II) using either α-cyano-hyroxycinnamic acid (CHCA) or sinapinic acid matrices. Each complex was either spotted directly on an H4 (hydrophobic surface) ProteinChip array or reduced with DTT at 100 °C for 1 min prior to spotting. One microliter of diluted sample was put onto the chip and allowed to dry. Because the sample was prepared using MALDI-type conditions, the normal washing step was bypassed. Under other conditions in which buffers incompatible with subsequent mass analysis (i.e., detergents, urea, salts) are present, the ProteinChip array is washed to remove interfering substances. Half a microliter of saturated matrix solution in 0.5% trifluoroacetic acid, 50% acetonitrile, and 49.5% water was then added to each spot and allowed to dry prior to mass analysis.

RESULTS

Transcription of Mammaglobin and Lipophilin B in Various Tissues. RT-PCR was used to examine the message levels of mammaglobin and lipophilin B in a number of breast tumors, other tissues, and tumor cell lines. These data (Figure 1) demonstrate a correlation between mRNA expression of lipophilin B and mammaglobin in the majority of breast tumors with a Spearman rank correlation coefficient (*34*) of 79% for 24 tumor samples ($p < 0.0001$). The same coefficient, calculated for another breast-linked marker, Her2/neu, for which there is no evidence of linkage to mammaglobin and lipophilin B, is 16% for the 24 samples ($p \leq 0.42$). This correlation is not as clear in normal tissues, where

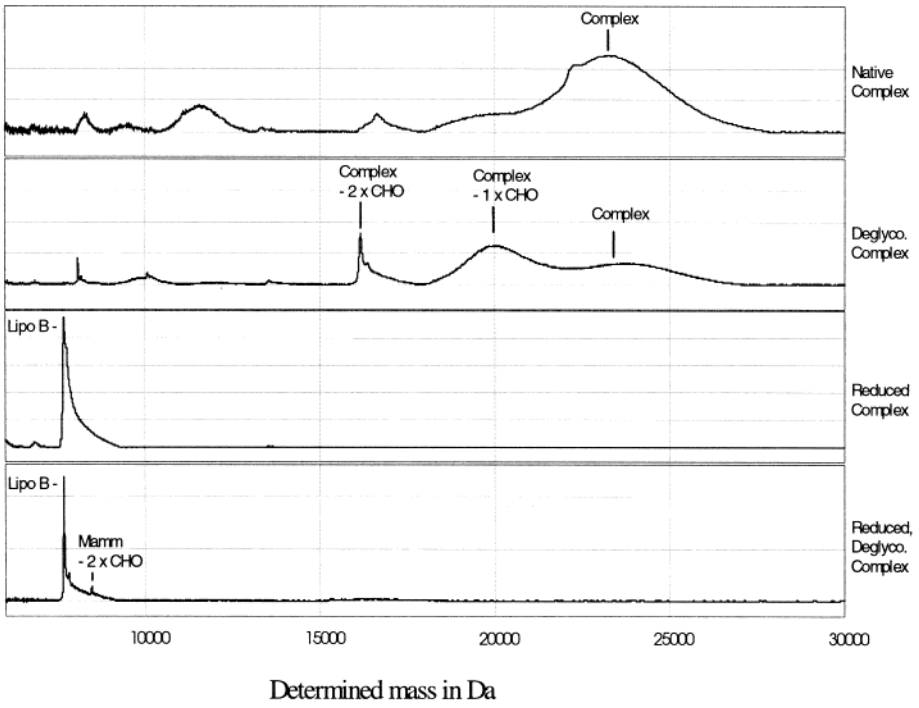


FIGURE 3: SELDI-TOF MS of the mammaglobin complex. Native complex; PNGase F treated complex; reduced native complex; and PNGase F treated, reduced complex were spotted onto a hydrophobic ProteinChip array and allowed to dry. Matrix was added and mass analysis performed in a Ciphergen PBS II externally calibrated with appropriate protein or peptide standards. Peak labels: complex, native complex with determined mass of 23,420; complex $-1 \times$ CHO, complex that has lost one carbohydrate, determined mass = 20,074; complex $-2 \times$ CHO, completely deglycosylated complex, determined mass = 16,173 (predicted for deglycosylated complex = 16,163); LipoB, lipophilin B, determined mass = 7686 (predicted mass = 7687); Mamm $-2 \times$ CHO, completely deglycosylated mammaglobin, determined mass = 8473 (predicted mass for deglycosylated mammaglobin = 8481). Glycosylated mammaglobin is not readily detected with any of the matrices tried; a small amount can be seen in the reduced native complex at a mass of 15,400. The novel peaks in the deglycosylated samples were not seen in a blank deglycosylation reaction.

lipophilin B had a less specific mRNA profile than mammaglobin with elevated mRNA levels in skeletal muscle, retina, and skin (Figure 1). Spearman rank correlation coefficients for the 26 tissue types tested are 42% ($p \leq 0.03$) for the mammaglobin−lipophilin B pairings and 28% ($p \leq 0.17$) for the lipophilin B−Her2/neu pairings.

Purification of the Complex. Mammaglobin/lipophilin B complex was purified by standard biochemical techniques to >90% purity as assayed by N-terminal sequencing of the purified liquid pool after reversed phase chromatography or by Coomassie Blue staining followed by densitometric analysis of the lipophilin B band (Figure 2A). The yield of complex is ~1 mg/L of conditioned medium. This purified complex was then used to create a monoclonal antibody to the native form of mammaglobin that does not bind to recombinant mammaglobin produced in *Escherichia coli* (data not shown). The availability of this antibody will facilitate future studies on the native form in vitro and in vivo and has allowed for the making of an affinity column for the complex, which simplifies the purification of native protein.

Signal Peptide Cleavage Site. N-Terminal sequencing of the purified protein in solution and of blots with the proteins either associated or reduced yields the following sequences (first residue trace is shown in Figure 2A):

lipophilin B: EFxPALVSEL...

mammaglobin: GSGxPLLENV...

(Cysteinyl residues are not seen using this method; "x" is interpreted as Cys.) These cleavage sites are consistent with the predicted sites reported in the deposited sequences in the Entrez protein sequence database (accession for mammaglobin, Q13296, and for lipophilin B, NP_006542).

Mammaglobin Glycosylation. Purified mammaglobin stains easily by a carbohydrate-specific stain (Figure 2B), whereas lipophilin B is not seen with this type of staining. On the other hand, lipophilin B is easily detected by Coomassie stain on a reduced SDS−PAGE sample, although there is no visible mammaglobin band prior to deglycosylation. The difficulty of seeing highly glycosylated proteins by standard staining methods has already been noted, and mammaglobin appears when stained according to a method designed to detect highly glycosylated proteins (*35*) (data not shown). There are two potential N-linked glycosylation sites in the mammaglobin sequence: one at Asn 35 (Asn-Ala-Thr) and one at Asn 50 (Asn-Gln-Thr). N-Terminal sequencing of the glycosylated complex can read the amino acids prior to and following Asn 35, but this amino acid is not seen despite the normally good yield for asparagines. This is interpreted as a modification on this Asn. N-Terminal sequencing by Edman degradation was not possible through amino acid 50 due to decreasing yields with each cycle. However, mass spectrometry (MS) of the native and partially deglycosylated mammaglobin gives strong evidence that both sites are modified. The native complex appears on SELDI-MS as a broad 23,420 Da peak with a 20 kDa shoulder consistent with the two forms of mammaglobin seen in cell culture supernatants by Western blotting. Treatment with PNGase F, an enzyme that cleaves N-linked sugars completely leaving

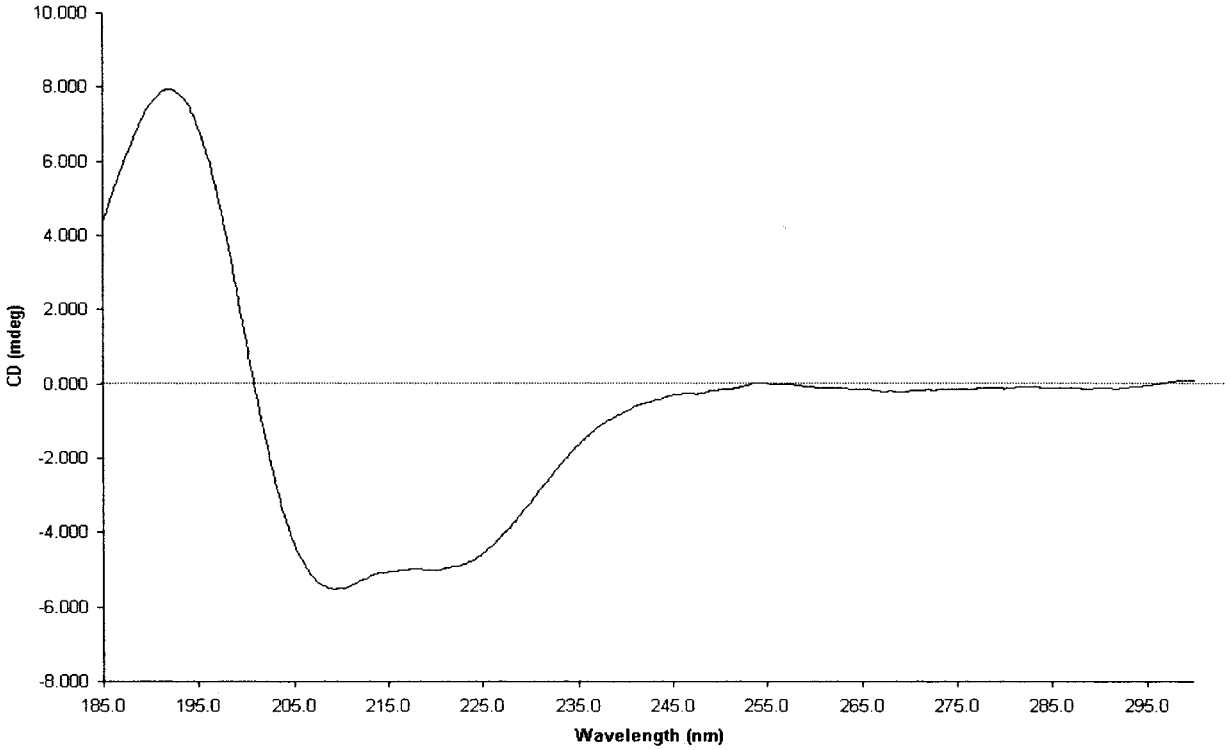
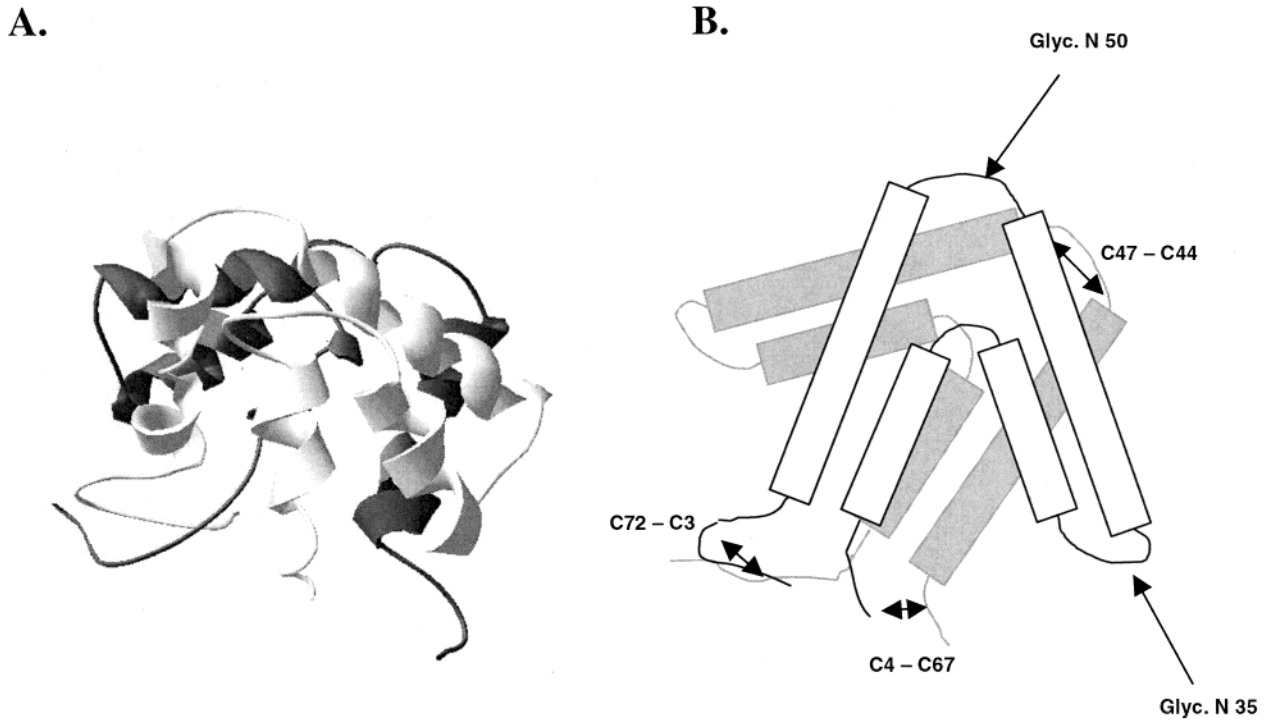


FIGURE 4: CD spectrum for mammaglobin complex. Mammaglobin complex was dialyzed against water and a CD spectrum collected from 185 to 300 nm on a JASCO J-810 CD spectrophotometer. Characteristic minima at 192 nm ($\pi \rightarrow \pi^*$), 209 nm ($\pi \rightarrow \pi^*$), and 220 nm ($n \rightarrow \pi^*$) were detected by inspection of first-derivative graphs of this spectrum and are indicative of the largely α-helical structure of the complex in solution.

the side chain only slightly modified (asparaginyls are converted to aspartyls in the process), increases the amount of the 20 kDa form, and a new peak is detected at 16,173 Da (Figure 3). Cleavage occurs in steps of 3346 and 3901 Da and yields a molecule with a molecular mass that is in close agreement with the predicted molecular mass of 16,163 Da. The broadness of the glycosylated peaks is due to the carbohydrates still attached to the protein and disappears in the deglycosylated protein (Figure 3: "complex $-2 \times$ CHO").

Mammaglobin−Lipophilin B Association. Mammaglobin and lipophilin B copurify through numerous different biochemical purification steps. They also comigrate on SDS−PAGE gels under nonreducing conditions. When blotted on PVDF membranes and sequenced, the diffuse band, which contains mammaglobin by Western blotting, also contains equimolar amounts of lipophilin B. This association of mammaglobin and lipophilin B can be broken by pretreatment with reducing agents such as 10 mM DTT, making an association by disulfide linkage likely. Prior to reduction, the center of the mammaglobin band is at ∼25 kDa on SDS−PAGE gels. After reduction, this band shifts downward to ∼20 kDa, consistent with a loss of one molecule of lipophilin B per complex (predicted molecular mass = 7.6 kDa). Below the 10 kDa molecular mass marker a new band appears upon reduction, which is absent in the nonreduced gels. This band can be stained using conventional silver staining and was revealed to be lipophilin B by N-terminal sequencing. Additionally, upon reduction of the native complex, a new peak is detected by MS with a molecular mass of 7686 Da, which closely matches the predicted molecular mass for a variant of lipophilin B (7687 Da). This variant is a result of a single nucleotide polymorphism in lipophilin B, which gives rise to a P to L mutation and is found in MDA-MB415 cells (*36*). The PNGase F treated complex gives rise to two peaks: a lipophilin B peak also detected in the untreated sample and new peak for deglycosylated mammaglobin at 8473 Da (Figure 3, bottom two panels).

Structure of the Mammaglobin/Lipophilin B Complex. CD spectroscopy of the complex yields a spectrum consistent with a largely α-helical protein. Characteristic peaks at 192, 209, and 220 nm are observed by inspecting the first-derivative graph of the spectrum shown in Figure 4. The amino acid sequences of mammaglobin and lipophilin B were aligned to uteroglobin and modeled on the basis of the structure of this molecule (Figure 5A). The model predicts a head to tail orientation of the two chains with each forming four α-helices (Figure 5B) and nicely aligns the cysteinyl residues from mammaglobin with those from lipophilin, predicting three disulfide bridges between the two. The head to tail orientation is further supported by ESI-MS data (Figure 6). A peptide matching the molecular mass of the combined N-terminal Cys-containing tryptic fragment of mammaglobin with the C-terminal Cys-containing tryptic fragment of lipophilin B is seen only in the nonreduced sample (Figure 6A). In the reduced run, both the free C terminus of lipophilin B (Figure 6B) and the N terminus of mammaglobin are seen (Figure 6C). The glycosylation sites are found in loops at each pole of the globule, allowing for attached carbohydrates to extend in opposite directions. Secondary structure prediction using a neural network based system is in close agreement with the predictions of the model for where helices will be found and further increases confidence in the molecular model.



C.

```
1. Mammaglobin

GSGCPLLENVISKTINPQVSKTEYKELLQEFIDDNATTNAIDELKECFLNQTDETLSNVEVFMQLIYDSSLCDLF

----HHHHHHHHHHH-HH---HHHHHHHHHH----H-HHHHHHHHHHHH---HHHHHHHHHHHHHH-------H-    Predicted

-------HHHHHHH--------HHHHHHHHH---HHHHHHHHHHHHHHH---HHHHHHHHHHHHHHH--------    Model

2. Lipophilin B

EFCPALVSELLDFFFISEPLFKLSLAKFDAPPEAVAAKLGVKRCTDQMSLQKRSLIAEVLVKILKKCSV

---HHHHHHHHHH-----HHHHHH------HHHHHHHHHH-HHHHHHH-HHHHHHHHHHHHHH--          Predicted

-------HHHHHHH-----HHHHHHHHH---HHHHHHHHHHHHHHH---HHHHHHHHHHHHHH-----          Model
```

FIGURE 5: Mammaglobin complex with lipophilin B: (A) Molecular model of the complex generated by SwissModel using the coordinates of rat uteroglobin (1UTR) as a template. The complex is predicted to be a globular, all α-helical molecule with a pocket for a steroid-like ligand. (B) Schematic of the molecular model. Double-headed arrows indicate cysteine bridges; Glyc. indicates the location of predicted N-linked glycosylation sites. In both panels A and B, mammaglobin is depicted in white, lipophilin B in gray. (C) Correlation of the model with predicted secondary structure. Helical structures in the model align well with those predicted by nnpredict from the amino acid sequence alone.

DISCUSSION

Although there have been numerous reports on the importance of mammaglobin message as a marker for breast cancer (*1*, *3*), the protein has not yet been well characterized—partially due to the difficulty of purifying sufficient quantities of the native form. We report the successful purification of the mammaglobin complex that has allowed us to generate an antibody that can recognize the native protein and discriminates between native and recombinant mammaglobin. This antibody has been used for immunoaffinity purification of the mammaglobin complex, enabling a simplified purification scheme for the molecule from various native sources.

Analyses of the purified molecule indicate that the assembly proceeds with cleavage of the signal peptides at the predicted locations for both mammaglobin and lipophilin B, glycosylation on mammaglobin alone, and disulfide formation between mammaglobin and lipophilin B. Uteroglobin family members tend to be homo- and heterodimers of either two uteroglobin-like molecules or a mammaglobin-like molecule bound to a lipophilin-B like molecule (*10*, *12*, *37*). Mammaglobin appears to follow this trend as well. It dimerizes with lipophilin B and is oriented in an antiparallel manner that allows for the formation of three disulfide bridges between the two molecules. Mammaglobin protein is also modified by the addition of two carbohydrate chains, one at each consensus Asn-linked glycosylation site. These chains account for ~30% of the total molecular mass of the complex, shifting the molecular mass from 16 kDa for the protein components to 23 kDa for the fully assembled complex. CD, molecular modeling, and secondary structure prediction on the amino acid sequence predict that this complex is largely α-helical with four helices per indepen-

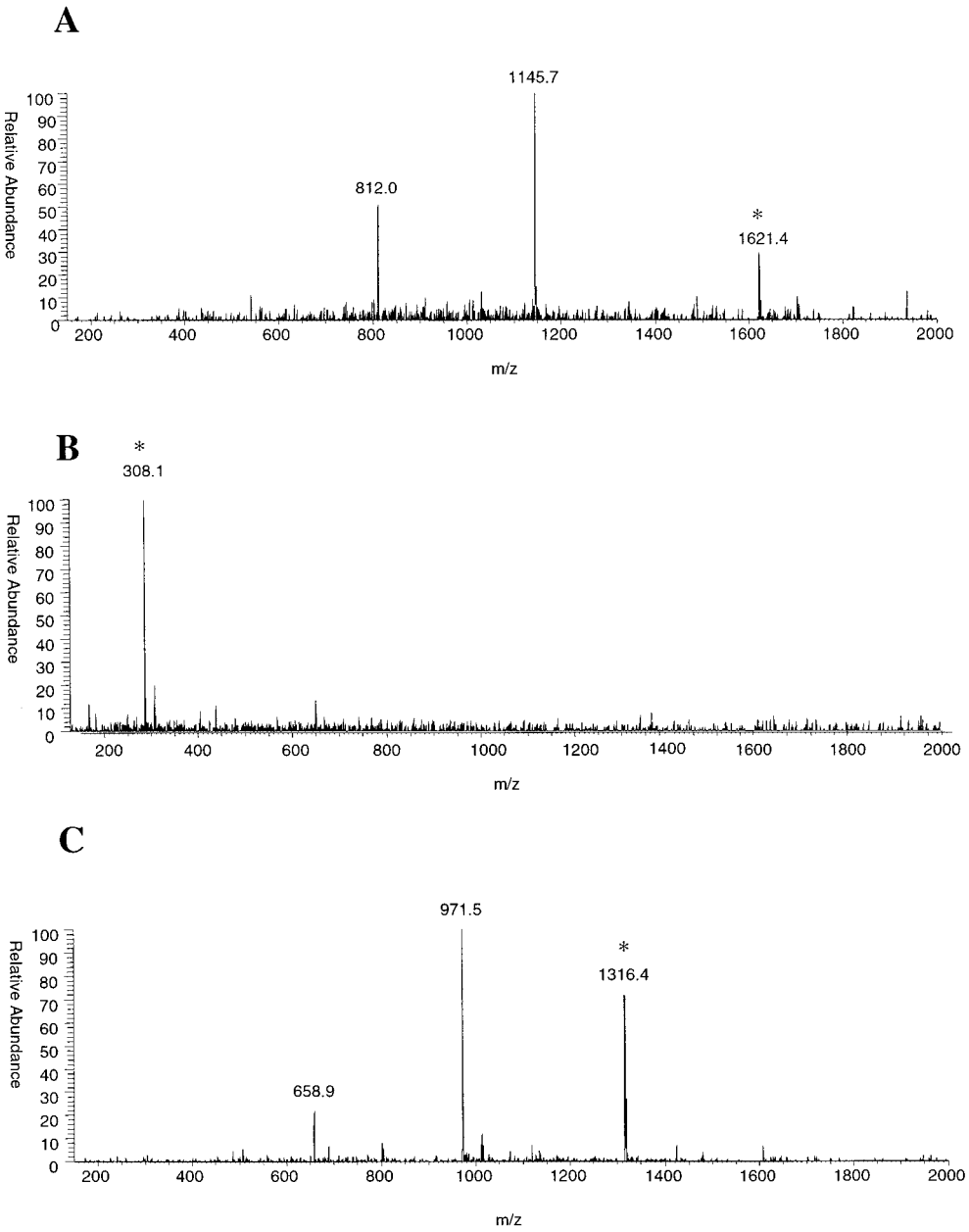


FIGURE 6: LC-ESI-MS spectra confirming the head to tail orientation of the complex: (A) Spectrum containing the disulfide-bonded N terminus of mammaglobin with the C terminus of lipophilin B. Upon tryptic digestion and reversed phase HPLC−MS the combined peptides for mammaglobin (GSGCPLLENVISK) and lipophilin B (CSV) are detected at 1621.4 Da (predicted for the bonded peptide = 1622.5 Da). Other peptides from the complex coelute. Peptides of interest in this figure are indicated with (∗). (B) Spectrum containing the free C-terminal fragment of lipophilin B. In reduced, digested complex free C-terminal peptide for lipophilin B is detected at 308.1 Da (predicted for the −CSV peptide = 308.1 Da). (C) Spectrum containing the free N-terminal fragment of mammaglobin. In reduced, digested complex free N-terminal peptide for mammaglobin is detected at 1316.4 Da (predicted for the GSGCPLLENVISK peptide = 1316.7 Da).

dent polypeptide chain. Assembly of the folded molecules forms a small α-helical globule and creates a hydrophobic pocket capable of binding steroid-like molecules in the reduced state. On loops between helices on either side of the globule the carbohydrates are attached, giving rise to a spherical molecule with a protruding carbohydrate chain on either end. The availability of a purification procedure for the native complex and increased understanding of its biochemical properties will allow investigation into the role this complex plays in cancer and will greatly facilitate the development of protein-based detection of mammaglobin. Using an ELISA-based format to quantify the serum levels of the complex in breast cancer patients may increase the selectivity of mammaglobin as a diagnostic and prognostic marker.

REFERENCES

1. Watson, M. A., and Fleming, T. P. (1996) *Cancer Res. 56*, 860−865.
2. Min, C. J., Tafra, L., and Verbanac, K. M. (1998) *Cancer Res. 58*, 4581−4584.
3. Watson, M. A., Dintzis, S., Darrow, C. M., Voss, L. E., DiPersio, J., Jensen, R., and Fleming, T. P. (1999) *Cancer Res. 59*, 3028−3031.
4. Houghton, R. L., Dillon, D. C., Molesh, D. A., Zehentner, B. K., Xu, J., Jiang, J., Schmidt, C., Frudakis, A., Repasky, E., Maltez, F. A., Nolasco, M., Badaro, R., Zhang, X., Roche, P.

C., Persing, D. H., and Reed, S. G. (2001) *Mol. Diagn. 6*, 79−91.
5. Colpitts, T. L., Billing, P., Granados, E., Hodges, S., Menhart, N., Russell, J., and Stroupe, S. (2000) *Ann. N. Y. Acad. Sci. 923*, 312−315.
6. Colpitts, T. L., Billing-Medel, P., Friedman, P., Granados, E. N., Hayden, M., Hodges, S., Menhart, N., Roberts, L., Russell, J., and Stroupe, S. D. (2001) *Biochemistry 40*, 11048−11059.
7. Lehrer, R. I., Xu, G., Abduragimov, A., Dinh, N. N., Qu, X. D., Martin, D., and Glasgow, B. J. (1998) *FEBS Lett. 432*, 163−167.
8. Zhao, C., Nguyen, T., Yusifov, T., Glasgow, B. J., and Lehrer, R. I. (1999) *Biochem. Biophys. Res. Commun. 256*, 147−155.
9. Becker, R. M., Darrow, C., Zimonjic, D. B., Popescu, N. C., Watson, M. A., and Fleming, T. P. (1998) *Genomics 54*, 70−78.
10. Morize, I., Surcouf, E., Vaney, M. C., Epelboin, Y., Buehner, M., Fridlansky, F., Milgrom, E., and Mornon, J. P. (1987) *J. Mol. Biol. 194*, 725−739.
11. Claessens, F., Rushmere, N. K., Davies, P., Celis, L., Peeters, B., and Rombauts, W. A. (1990) *Mol. Cell. Endocrinol. 74*, 203−212.
12. Hard, T., Barnes, H. J., Larsson, C., Gustafsson, J. A., and Lund, J. (1995) *Nat. Struct. Biol. 2*, 983−989.
13. Bochskanl, R., Wirth, B., and Kirchner, C. (1988) *Hum. Reprod. 3*, 844−850.
14. Maccioni, M., Riera, C. M., and Rivero, V. E. (2001) *J. Reprod. Immunol. 50*, 133−149.
15. Maccioni, M., Rivero, V. E., and Riera, C. M. (1998) *Clin. Exp. Immunol. 112*, 159−165.
16. Moreno, J. J. (1997) *Gen. Pharmacol. 28*, 23−26.
17. Mukherjee, A. B., Kundu, G. C., Mantile-Selvaggi, G., Yuan, C. J., Mandal, A. K., Chattopadhyay, S., Zheng, F., Pattabiraman, N., and Zhang, Z. (1999) *Cell Mol. Life Sci. 55*, 771−787.
18. Peri, A., Cordella-Miele, E., Miele, L., and Mukherjee, A. B. (1993) *J. Clin. Invest. 92*, 2099−2109.
19. Andersson, O., Nordlund-Moller, L., Barnes, H. J., and Lund, J. (1994) *J. Biol. Chem. 269*, 19081−19087.
20. Dierynck, I., Bernard, A., Roels, H., and De Ley, M. (1996) *Mult. Scler. 1*, 385−387.
21. Holland, P. M., Abramson, R. D., Watson, R., and Gelfand, D. H. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 7276−7280.
22. Schneeberger, C., Speiser, P., Kury, F., and Zeillinger, R. (1995) *PCR Methods Appl. 4*, 234−238.
23. Babcook, J. S., Leslie, K. B., Olsen, O. A., Salmon, R. A., and Schrader, J. W. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 7843−7848.
24. Laemmli, U. K. (1970) *Nature 227*, 680−685.
25. Blum, H., Beier, H., and Gross, H. J. (1987) *Electrophoresis 8*, 93−99.
26. Thornton, D. J., et al. (1994) *Methods Mol. Biol. 32*, 119−128.
27. Matsudaira, P. (1987) *J. Biol. Chem. 262*, 10035−10038.
28. Guex, N., Diemand, A., and Peitsch, M. C. (1999) *Trends Biochem. Sci. 24*, 364−367.
29. Hard, T., Barnes, H. J., Larsson, C., Gustafsson, J. A., and Lund, J. (1995) *Nat. Struct. Biol. 2*, 983−989.
30. Hein, J. (1990) *Methods Enzymol. 183*, 626−645.
31. Frishman, D., and Argos, P. (1995) *Proteins 23*, 566−579.
32. Kneller, D. G., Cohen, F. E., and Langridge, R. (1990) *J. Mol. Biol. 214*, 171−182.
33. McClelland, J. L., and Rumelhart, D. E. (1988) *Explorations in Parallel Distributed Processing*, MIT Press, Cambridge, MA.
34. Rosner, B. (1990) in *Fundamentals of Biostatistics* (Payne, M. R., Hankinson, S., and London, S., 3rd, Eds.), pp 451−453, PWS-KENT Publishing.
35. Gradilone, S. A., Arranz, S. E., and Cabada, M. O. (1998) *Anal. Biochem. 261*, 224−227.
36. Personal communication (2001).
37. Parker, M. G., White, R., Hurst, H., Needham, M., and Tilly, R. (1983) *J. Biol. Chem 258*, 12−15.

BI0159884