Nanodiamond for Sample Preparation in Proteomics
Felipe Perona Martinez, Andreas Nagl, Sona Guluzade, and Romana Schirhagl*

Groningen University, University Medical Center Groningen, Antonius Deusinglaan 1, 9713 AW Groningen, The Netherlands

ABSTRACT: Protein analysis of potential disease markers in blood is complicated by the fact that proteins in plasma show very different abundances. As a result, high-abundance proteins dominate the analysis, which often render the analysis of low-abundance proteins impossible. Depleting high-abundance proteins is one strategy to solve this problem. Here, we present, for the first time, a very simple approach based on selective binding of serum proteins to the surface of nanodiamonds. In our first proof-of-principle experiments, we were able to detect, on average, eight proteins that are present at a concentration of 1 ng/mL (instead of 0.5 ng/mL in the control without sample preparation). Remarkably, we detect proteins down to a concentration of 400 pg/mL after only one simple depletion step. Among the proteins we could analyze are also numerous disease biomarkers, including markers for multiple cancer forms, cardiovascular diseases, or Alzheimer’s disease. Remarkably, many of the biomarkers we find also could not be detected with a state-of-the-art ultrahigh-performance liquid chromatography column (which depletes the 64 most-abundant serum proteins).

It is believed that the majority of disease markers are still unidentified, since they are among the low-abundance proteins in plasma. However, recently, several methods have been developed to deplete high-abundance proteins from serum, thus allowing the analysis of low-abundance proteins. For instance, there are commercially available high-pressure liquid chromatography (HPLC) columns, which contain antibodies against high-abundance proteins and thus retain them in the column. While initially only a few proteins were depleted, new columns are available that deplete several tens of proteins simultaneously. Another approach is to use nanoparticles, which bind to certain proteins. For instance, Liu et al. used several steps of precipitation with polyethylene glycol (PEG) for this purpose, followed by depletion with one of the above-mentioned antibody columns. Large amounts of proteins have also been identified. However, the authors used more-complex multistep protocols. One alternative that does not require specific antibodies is represented by molecularly imprinted polymer particles. To produce these, one must imprint a polymer with the proteins that need to be depleted. However, in order to achieve this, one must know the proteins that should be depleted and have them available. This issue was solved elegantly by Yang et al., who imprinted with the full bovine serum. By varying the concentration that was used for imprinting, they could tune the amount of proteins that are adsorbed.

An alternative approach for protein enrichment is combinatorial peptide ligand libraries (CPLLs). To produce the library, beads are coated with many different covalently attached peptides. These bind different proteins in the serum, which are thus removed from the sample. The remaining serum is strongly depleted of all types of proteins, including the most abundant ones. This approach does not require specific antibodies or prior knowledge and has already been successfully applied to several different samples with a complex proteome.

However, despite these efforts, the depletion of high-abundance proteins still remains an issue. Here, we show a simple, fast, and cost-effective method to achieve high-abundance protein depletion. To achieve protein depletion, we use the fact that only some proteins bind to nanodiamonds. Our approach works similarly to CPLLs in the sense that there are particles that bind to many different proteins. However, we have the advantage that our particles are slightly simpler and, since there are no biomolecules attached, they are likely more durable. A disadvantage is probably that the surface chemistry is less complex and, thus, probably binds less proteins than the complex surface of CPLLs.

The nanodiamonds in our experiments have traditionally been used as abrasive and are thus readily available commercially. They also recently gained popularity for their magneto-optical properties and their use as long-term fluorescent labels as well as their use in drug delivery. However, their application in depleting high-abundance proteins from plasma is entirely new.

MATERIALS AND METHODS

To eliminate high-abundance proteins, nanodiamonds and NaCl were added to the serum. As a result, aggregates precipitate. Since several of the high-abundance proteins bind poorly to the diamond surface, one can deplete them by

Received: March 22, 2019
Accepted: July 10, 2019
Published: July 10, 2019
removing the supernatant. When the protein corona on the diamond surface is analyzed via mass spectrometry, we find an increased number of low-abundance proteins. For a schematic representation of the protein depletion, see Figure 1.

**Sample Preparation with Carbon Black.** Next, we answered whether the protein depletion is specific for diamond nanoparticles. To this end, we prepared our samples in exactly the same way as with FND, except the FNDs was replaced with carbon black.

**Protein Analysis.** The samples were analyzed via nanoLC−MS/MS on an Ultimate 3000 system (Dionex, Amsterdam, The Netherlands) interfaced online with a Q-ExactivePlus (Orbitrap) mass spectrometer (Thermo Fisher Scientific Inc., Waltham, MA, USA). Peptide mixtures were loaded onto a 5 mm × 300 μm i.d. trapping microcolumn that was packed with C18 PepMAP100 5 μm particles (Dionex) in 2% acetonitrile in 0.1% formic acid at the flow rate of 20 μL/min. After loading and washing for 3 min, peptides were backflush eluted onto a 15 cm × 75 μm i.d. nanocolumn and packed with C18 PepMAP100 1.8 μm particles (Dionex). The following mobile phase gradient (total run time: 75 min) was delivered at the flow rate of 300 nL/min: 2%−50% of solvent B in 60 min; 50%−90% B in 1 min; 90% B during 13 min, and back to 2% B in 1 min (held for 15 min). Solvent A was 100:0 H2O/acetonitrile (v/v) with 0.1% formic acid, and solvent B was 0:100 H2O/acetonitrile (v/v) with 0.1% formic acid. Peptides were infused into the mass spectrometer via dynamic nanospray probe (Thermo Fisher Scientific, Inc.) with a stainless steel emitter (Thermo Fisher Scientific, Inc.). The typical spray voltage was 1.8 kV with no sheath and auxiliary gas flow; ion transfer tube temperature was 275 °C. Mass spectrometer was operated in data-dependent mode. DDA cycle consisted of the survey scan within m/z 300−1650 at the Orbitrap analyzer with target mass resolution of 70 000 (full width at half-maximum (fwhm) at m/z 200), followed by MS/MS fragmentations of the top 10 precursor ions. Singly charged ions were excluded from MS/MS experiments and m/z of fragmented precursor ions were dynamically excluded for an additional 20 s.

**Data Processing.** The software PEAKS Studio version 7 (Bioinformatics Solutions, Inc., Waterloo, Canada) was applied to the spectra generated by the Q-Exactive Plus mass spectrometer to search against the protein sequence database UniProtKB/Trembl of the UniProt Knowledgebase (UniProtKB), limited to protein sequences of *Homo sapiens* (a search including the entire database was performed as well, to rule out the relevance of possible contaminations). Searching for the fixed modification carbamidomethylation of cysteine and the variable post-translational modifications oxidation of methionine was done with a maximum of five post-translational modifications per peptide at a parent mass error tolerance of 10 ppm and a fragment mass tolerance of 0.02 Da. The false discovery rate was set at 0.1%.

From the mass spectrometry, one obtains spectral counts. These reflect how often protein fragments are found that can be attributed to a certain protein. However, larger proteins naturally lead to more fragments. To compensate for this fact, one must calculate normalized spectral counts. These give a semiquantitative measure for the (relative) the concentration of a certain protein in the sample. The normalized spectral counts are calculated by using the following equation:25−50

\[
\text{NpSpC}_k = 100 \times \frac{(\text{SpC}/\text{MW})_k}{\sum_{i=1}^{n} (\text{SpC}/\text{MW})_i}
\]

where NpSpC$_k$ is the normalized percentage of spectral count (which is the number of spectra associated with a protein) for
protein $k$, $SpC$ is the spectral count identified, and $MW$ is the molecular weight (in daltons) of the protein $k$.

Waterfall plots were created by comparing the protein lists with the human proteome project database (HPP-DB). The concentrations in the database reflect the current knowledge from selected references.

**RESULTS**

When we precipitate proteins together with nanodiamonds in a salt-containing medium, we find that some of the most abundant serum proteins bind poorly to the nanodiamonds surface. We then used liquid chromatography coupled with mass spectrometry (LCMS) analysis to determine which proteins can be found on the diamond surface. We typically find several hundred proteins on our diamond surface. Figure 2 summarizes the depletion that we find for different media.

To determine the optimal conditions where most low-abundance proteins bind to the surface while high-abundance proteins remain in the supernatant, we tested different salt concentrations. The concentrations that we chose were near the physiological concentration of 6.9 mg/mL NaCl. In addition to varying the salt concentration, we also investigated the effect of washing in order to differentiate between the hard and soft protein corona. The soft corona (before washing) contains loosely and strongly binding proteins. The hard corona is what remains after washing and only contains strongly binding proteins. For most cases, we do see a small decrease in high-abundance proteins after washing. In addition to quantifying the most abundant proteins, we were also interested in the composition of the protein corona. Figure 3 shows which categories of proteins we find on which sample.

Figure 2. Depletion of high-abundance proteins with nanodiamonds. Compared to the control (serum without any treatment), shown in green, the amount of high-abundance proteins that is found by mass spectrometry is significantly reduced when these were previously depleted with nanodiamonds. Different media are used to precipitate protein-coated diamonds, and the depletion is compared. Error bars are generated from three different independent experiments and represent the standard error of the mean.

To make this classification, we ranked the proteins from the highest concentration to the lowest concentration. We took into account all proteins in the top 50%. We chose to use the top 50% here, since, for lower-abundance proteins, these classifications are scarce or not available at all. The groups, based on biological functions that we could distinguish, are apolipoproteins (APO), complement factors (COM), other O, immunoglobulins (IG), acute phase proteins (ACP), and coagulation factors (COA). We found large differences in the corona composition. Whereas, in the control, the top 50% of the corona consists of apolipoproteins, the diamond samples are more diverse. Most likely binding to the diamond surface occurs via electronegative oxygen groups on the diamond surface, which can interact with electropositive groups within proteins. While we could not establish a clear relationship between, for instance, binding and the isoelectric point of the proteins, we do often see proteins binding whose function in biology is to bind to electronegative structures. What we observe is similar to CPLLs, which offer a rich surface chemistry, to which proteins can bind. Similar to CPLLs, we also do not target a specific protein or a number of protein (as an antibody column does) but rather deplete anything that does not bind. Next, we compared the samples based on their ability to detect low-abundance proteins. To this end, we used so-called “waterfall plots”. To construct a waterfall plot, the
protein lists are compared with the database. Figure 4 shows one of these waterfall plots, which we obtained for the best condition (serum +6.9 mg/mL NaCl + FND). The proteins in the database are plotted in order of decreasing concentrations. Every protein that is identified in the sample receives a blue dot. To illustrate the improvement, a dotted line is used to indicate the lowest concentrated protein that we could detect with the control. The proteins below that dotted line (marked with a rectangle) are only accessible with the diamond sample preparation step.

Most interesting for proteomics are proteins with concentrations of <1 ng/mL. These are challenging to analyze without specialized sample preparation. In Figure S, we compare how many of these low-abundance proteins one can find with each sample preparation method. The condition with serum +6.9 mg/mL NaCl + FND, which can reveal eight proteins, on average, gives the best results. For instance, the control only gives 0.5 proteins, on average.

As a final assessment of usefulness of our method, we compared the proteins that we could identify with proteins that are already used as biomarkers in the literature. Table 1 gives a few examples, which seemed to be most interesting to us.

### Table 1. Examples of Proteins Identified in the Best Sample (Serum +6.9 mg/mL NaCl + FND) That Could Be Detected Neither in the Reference nor with a State-of-the-Art Depletion Column with 64 Antibodies and Their Clinical Relevance

<table>
<thead>
<tr>
<th>Protein</th>
<th>Clinical Relevance</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>von Willebrand factor</td>
<td>Willebrand disease, the most common inherited bleeding disorder</td>
<td>31</td>
</tr>
<tr>
<td>Tetrantecin</td>
<td>marker for disease activity in patients with rheumatoid arthritis</td>
<td>32</td>
</tr>
<tr>
<td>Proteoglycan 4</td>
<td>diagnostic biomarker for COPD (chronic obstructive pulmonary disease)</td>
<td>33</td>
</tr>
<tr>
<td>Vitamin D-binding protein</td>
<td>risk factor for colorectal cancer</td>
<td>34</td>
</tr>
<tr>
<td>Fibulin-1</td>
<td>cardiovascular risk markers in chronic kidney disease and diabetes</td>
<td>35</td>
</tr>
<tr>
<td>Hornerin</td>
<td>aberrantly expressed in breast cancer</td>
<td>36</td>
</tr>
<tr>
<td>Hepatocyte growth factor activator</td>
<td>diagnostic value for numerous diseases, as well as age and pregnancy</td>
<td>37</td>
</tr>
<tr>
<td>Apolipoprotein M</td>
<td>suspected to be a biomarker for certain diabetes types</td>
<td>38</td>
</tr>
<tr>
<td>Endostatin</td>
<td>diagnosing malignant pleural effusions, antiangiogenic agent</td>
<td>39</td>
</tr>
<tr>
<td>Suprabasin</td>
<td>tumor endothelial cell marker</td>
<td>40</td>
</tr>
<tr>
<td>Angiogenin</td>
<td>used in prediction of failure on long-term treatment response and for poor overall survival in non-Hodgkin lymphoma (a certain cancer type)</td>
<td>41</td>
</tr>
<tr>
<td>Desmoplakin</td>
<td>biomarker for Creutzfeldt–Jakob disease</td>
<td>42</td>
</tr>
<tr>
<td>Ribonuclease 4</td>
<td>diagnosis of pancreatic cancer</td>
<td>43</td>
</tr>
</tbody>
</table>

#### Sample Preparation with Carbon Black

Finally, we wanted to determine if the depletion effect that we see is specific for diamond. When diamond is replaced in the above-mentioned experiments, as shown in Figure 6, we do not observe any depletion effects under any conditions. This finding indicates that the depletion of high-abundance serum proteins is indeed a peculiarity of diamond nanoparticles (or particles that resemble them). The main difference between carbon black and HPHT diamond is the content of SP2 vs SP3. While carbon black contains large amounts of SP2 (carbon black is actually more similar to graphite than it is to diamond), HPHT diamond is almost exclusively SP3 carbon. The consequence is that carbon black can interact with proteins via $\pi-\pi$ interactions (which are not available in diamond). If such groups are exposed on the protein surface, they will interact more with carbon black. Oxygen-containing polar groups, on the other hand, are more prominent on the diamond surface. Since graphitic layers are (apart from defects) saturated and give less opportunities for oxygen-containing groups.

#### CONCLUSIONS

While antibody-based depletion columns are generally quite expensive, nanodiamonds are surprisingly inexpensive, since they are commercially available mass products, which are used as abrasives. In addition, the depletion process is just one fast and straightforward step. While antibodies bind very specifically to a predefined target, here, we use a less-specific
approach. We believe that proteins bind to specific groups on the diamond. Diamond particles provide a rich surface chemistry, which provide all types of oxygen-containing groups that (similar to a CPLL) can interact with different proteins. During our experiments, we were able to deplete high-abundance proteins significantly. As a result, we have access to low-abundance proteins for analysis, which would otherwise be undetectable. With this simple method, we were able to detect proteins down to the pg/mL range. The best results (the highest number of low-abundance proteins) that we can achieve were found when salt was added in physiological concentrations. With this approach, we are able to detect several disease biomarkers, including, among others, markers for several cancer types, cardiovascular diseases, or kidney function.

**AUTHOR INFORMATION**

**Corresponding Author**

*E-mail: romana.schirhagl@gmail.com.*

**ORCID**

Romana Schirhagl: 0000-0002-8749-1054

**Author Contributions**

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

**Funding**

R.S. acknowledges financial support from FOM via Projectruimte Grant No. 1SPR3229 and for an ERC Starting Grant No. 714289—Stress Imaging. F.P.M. acknowledges support from the Chilean government via a CONICYT scholarship (Grant No. 72160222).

**Notes**

The authors declare no competing financial interest.

**REFERENCES**


**ACKNOWLEDGMENTS**

Furthermore, we would like to acknowledge support with mass spectrometry and sample preparation from Prof. Bischoff and Dr. Permentier and their team.

**ABBREVIATIONS**

APO, apolipoproteins; COM, complement factors; O, other; IG, immunoglobulins; ACP, acute phase proteins; COA, coagulation factors

![Figure 6. Comparison with carbon black. Compared with the control (green, just serum), we do not observe any significant depletion for any conditions using carbon black (blue). Also, the washing step did not improve the situation (red). We added the FND samples for comparison and for a positive control.](image-url)