Less is More: Membrane Protein Digestion Beyond Urea–Trypsin Solution for Next-level Proteomics*\textsuperscript{§}

Xi Zhang†\textsuperscript{§§}

The goal of next-level bottom-up membrane proteomics is protein function investigation, via high-coverage high-throughput peptide-centric quantitation of expression, modifications and dynamic structures at systems scale. Yet efficient digestion of mammalian membrane proteins presents a daunting barrier, and prevalent day-long urea–trypsin in-solution digestion proved insufficient to reach this goal. Many efforts contributed incremental advances over past years, but involved protein denaturation that disconnected measurement from functional states. Beyond denaturation, the recent discovery of structure/proteomics omni-compatible detergent n-dodecyl-β-D-maltopyranoside, combined with pepsin and PNGase F columns, enabled breakthroughs in membrane protein digestion: a 2010 DDM-low-TCEP (DLT) method for H/D-exchange (HDX) using human G protein-coupled receptor, and a 2015 flow/detergent-facilitated protease and de-PTM digestions (FDD) for integrative deep sequencing and quantitation using full-length human ion channel complex. Distinguishing protein solubilization from denaturation, protease digestion reliability from theoretical specificity, and reduction from alkylation, these methods shifted day(s)-long paradigms into minutes, and afforded fully automatable (HDX)-protein-peptide-(tandem mass tag)-HPLC pipelines to instantly measure functional proteins at deep coverage, high peptide reproducibility, low artifacts and minimal leakage. Promoting—not destroying—structures and activities harnessed membrane proteins for the next-level streamlined functional proteomics. This review analyzes recent advances in membrane protein digestion methods and highlights critical discoveries for future proteomics. Molecular & Cellular Proteomics 14: 10.1074/mcp.R114.042572, 2441–2453, 2015.

The goal of proteomics has grown out of matured protein identification toward the next level of function discovery, particularly including membrane proteins, through quantitative and structural proteomics at high coverage, high throughput and systems scale. Bottom-up membrane proteomics provides two paths to spearhead toward this goal. Path 1, to identify and quantify, at higher accuracy, proteins present in a mixture based on sampling fragments of each (modification-bearing or not), often aiming to find protein candidates for new biomarkers or drug targets. Path 1 represents the predominant pursuit in current membrane and global proteomics research and reviews. Path 2, to define, at atomic precision, the complete molecular identities of signaling-pivot membrane proteins—that are established high-priority therapeutic targets—under functional states and to comprehensively quantify their dynamic changes upon stimuli (Fig. 1). Path 2 includes concrete quantitation of pan-post-translational modification (PTM)\textsuperscript{1} percent site occupancy and structural labels such as H/D-exchange (HDX), both requiring reproducible peptides that cover near-full protein sequence. This approach complements crystallography, electron microscopy (EM), NMR, and top-down MS, can discover critical details and landscapes beyond current reach, yet is largely untapped—for technical hurdles (1, 2), thus this review focuses on Path 2. As downstream state-of-the-art HPLC and mass spectrometry technologies grow mature, generic, and widely accessible, digestion sample preparation methods for membrane proteins increasingly delimit the capacity of bottom-up proteomics.

\textsuperscript{1} The abbreviations used are: PTM, post-translational modification; HDX, H/D exchange; TM, transmembrane; DDM, n-dodecyl-β-D-maltopyranoside; TCEP, Tris-2-carboxyethylphosphine; TMT, tandem mass tag; DLT, DDM-low-TCEP digestion for HDX; FDD, flow/detergent-facilitated protease and de-PTM digestions for deep sequencing; h, human; LGIC, ligand-gated ion channel; GPCR, G protein-coupled receptor; GABA\textsubscript{A}, type A receptor; β\textsubscript{2}-AR, beta-2 adrenergic receptor; HEK293, human embryonic kidney cell line 293; PSM, peptide spectrum match; FDR, false discovery rate; C-C, disulfide-bonded Cys-loop; Mox, Methionine oxidation; RP, reversed-phase; UPLC, ultra-performance liquid chromatography; FASP, filter-aided sample preparation; dial-filtration, di-alysis-filtration; SP3, single-pot solid-phase-enhanced sample preparation; EM, electron microscopy; amphipols, amphipolymers; MSP, membrane scaffold proteins; PFG, PNGase F; cmc, critical micelle concentration; AN, aggregation number; CHS, cholesteryl hemisuccinate; MNG, maltose-neopentyl glycol; FA, steroid-based facial amphiphiles; BP, beta-strand peptide; DM, n-decyl-β-D-maltopyranoside; OG, n-octyl-β-D-glucoside.

From the †Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Boston, Massachusetts; ‡Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts

Received June 30, 2014, and in revised form, April 9, 2015
Published, MCP Papers in Press, June 16, 2015, DOI 10.1074/mcp.R114.042572
Author Contributions: X.Z. conceived and wrote paper.
Abundant in human genome and pivotal in cell signaling, transmembrane (TM) proteins are coveted therapeutic targets (3), yet tremendously difficult to study at all levels (1)—including proteomics (2, 4, 5). Nearly 25% of 29,375 unique protein sequences in human proteome contain one or more TM helices (3). Prominently, TM neurotransmitter receptors such as G protein-coupled receptors (GPCR) and ligand-gated ion channels (LGIC)—with most members yet to tap—exceed 50% of current therapeutic targets (3). However, hydrophobic and prone to aggregation, membrane proteins have long vexed bottom-up proteomics with under-representation in every metric, including: sequence coverage, peptide spectrum matches (PSMs), unique peptides, and peptide reproducibility (2, 4, 5)—regardless of global lysates or highly purified samples—contrasting facile soluble proteins.

Aggregation, though originated from hydrophobic TM domains, engages entire proteins and diminishes overall protease access, thus the chances to capture non-TM peptides at authentic quantity from aggregated samples are proven slim and unpredictable, and this fact cannot be reversed by simply changing database size. Global analysis without the fair share of membrane proteins is hardly global, most readily detectable soluble proteins—represented by two or more unique peptides—between independent digestions and sample states was routinely achieved in shotgun proteomics, until recently (6, 7), comprehensively reproducing peptide forms and abundances, which matters for peptide-centric peak-area or tandem mass tag (TMT)-based quantitation of expression, modification, and labels, remained out of reach (2, 4, 5).

The challenges in detecting peptides for membrane proteins are twofold: (1) forming peptides, and (2) detecting formed peptides. Using prevalent overnight urea–trypsin in-solution digestion, extreme three-dimensional 8-day HPLC separations benchmarked sequence coverage at 20–30% for membrane proteins (8), demonstrating deficient peptide formation pre-HPLC is a major unmet need, although many recent efforts to delve deeper focused on HPLC, mass spectrometer, and tandem MS strategies downstream to digestion (8, 9). Further, unaddressed PTMs of mammalian membrane proteins (such as extensive glycosylation at high site occupancy (10)), may prevent sequence identification, even if peptides are formed. Both hydrophobic TM domains and intricate PTMs, hallmarks of human signaling TM proteins, are important for functions. Detergents, inherited from protein extraction (typically SDS and Triton X-100), severely challenged downstream digestion and/or HPLC steps in global proteomics.
The “urea–trypsin solution” tradition contains three elements—urea denaturation, trypsin, and in-solution format—that each and all, deserve critical thinking for membrane proteins. Myriad resorts have been proposed to address membrane protein gel-free digestions for various needs over past decades (11, 12), yet attempts to denature and subdue proved ineffective. This review will focus on advances since 2009 for prospects to support the next-level streamlined high-coverage, high-throughput quantitative membrane proteomics (Table I).

### Table I

<table>
<thead>
<tr>
<th>Procedures and metrics</th>
<th>Traditional</th>
<th>On-filter in-solution</th>
<th>Protein on-bead</th>
<th>Pepsin beads</th>
<th>Enzyme column reactors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell lysis/protein extraction</td>
<td>SDS or Trition X-100</td>
<td>SDS</td>
<td>deoxycholate Na</td>
<td>SDS</td>
<td>CHAPS to MSP nanodiscs</td>
</tr>
<tr>
<td>Protein precipitation to remove detergent</td>
<td>urea</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>CHAPS/CHS</td>
</tr>
<tr>
<td>Protein pellet resolubilization</td>
<td>urea</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>DOM/CHS, slow addition</td>
</tr>
<tr>
<td>In-solution detergent exchange</td>
<td>--</td>
<td>urea diafiltration</td>
<td>urea protein on-bead</td>
<td>SDS removal by organic solvents; protein on-bead</td>
<td></td>
</tr>
<tr>
<td>Protein reduction and alkylaation</td>
<td>DTT, iodoaa</td>
<td>DTT, iodoaa</td>
<td>TCEP, Cl-acetamide</td>
<td>DTT, iodoaa</td>
<td>DTT, iodoaa</td>
</tr>
<tr>
<td>Digestion</td>
<td>ovnt(s)</td>
<td>ovnt(s)</td>
<td>ovnt; ethyl acetate extraction</td>
<td>ovnt (14 h)</td>
<td>offline, 1.5 M guanidinio_500 mM TCEP, 5 min</td>
</tr>
<tr>
<td>Peptide collection</td>
<td>--</td>
<td>filter centrifuge</td>
<td>C18 st, centrifuge</td>
<td>centrifuge</td>
<td>magnet DMSO</td>
</tr>
<tr>
<td>Peptide desalting</td>
<td>C18 st</td>
<td>C18 st</td>
<td>C18 st</td>
<td>_</td>
<td>online trap</td>
</tr>
<tr>
<td>Peptide drying/concentrating</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>Peptide labeling, reduction and alkylaation</td>
<td>TMT (optional)</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>TMT (optional); NH2OH (optional); iodoaa, 15 min</td>
</tr>
<tr>
<td>Post-label cleanup</td>
<td>yes (optional)</td>
<td>--</td>
<td>--</td>
<td>yes (optional) magnetic beads</td>
<td>--</td>
</tr>
</tbody>
</table>

**Ready for HPLC injection**

- Total protein-peptide time: > 15 h high, > 15 h or 19 h low, > 14.5 h low, > 40 h low, > 14 h low, > 6 min high from MSP, < 1 min < 9 min + 15 min, ~ 0
- Sequence coverage:
  - Peptide reproducibility: low med med med med med-high med-high (42%-45% a)
  - N-glyco PTM addressed?: no; no; yes no yes no no yes, fully
  - Automatable?: low low partial low partial partial full full
  - Protein activity pre-digestion: dead dead dead dead active active active

**Digestion competency**

- Path 1: id qt proteins in mixtures: low med-high med-high med-high med-high likely med likely med-high likely high
- Path 2: define full protein PTMs and HDX dynamics: low; no likely med; no likely med; no likely med-high; no likely med-high; no med med-high ultra high

---

a Sequence coverage also depends on HPLC method and sequence database used for search.
b Purified protein searched against specified sequence.
c Identified by searching against human proteome (original, downloaded May, 23, 2013 from Ensembl) at peptide FDR < 1%, using SEQUEST-HT and Percolator in PD 1.4. Signaling peptide sequence was not removed (a “59%” correction, assuming a 20-residue signaling peptide for a total 500-residue protein). MSP, membrane scaffold protein; FDD, flow/detergent-facilitated protease and de-PTM digestions for deep sequencing; st, StageTip; iodoaa, iodoacetamide; ovnt, overnight (14 h); N-glyco, N-glycosylation; med, medium; id, identify; qt, quantify; hp, human proteome search.
Real-time Membrane Protein Digestion for Future Proteomics

remove SDS, resolubilization in 8 M urea, DTT reduction, iodoacetamide alkylation, in-solution digestion with trypsin in 2 M urea overnight, peptide desalting with C18 to remove the high urea, peptide drying to remove C18 elution solvent and concentrate (often causing oxidation), and redissolving in HPLC sample buffer. Pronounced problems are: serious unpredictable protein loss to aggregation (14), high oxidation artifacts and contaminations introduced by laborious processing (15), biases over protease specificity and abundance (13, 16), low digestion efficiency, coverage, and peptide reproducibility (2, 4).

Previous efforts to improve membrane protein digestion have focused on proteases with strict specificity, such as trypsin, Lys-C, Glu-C, Glu-N, and Asp-N, aiming to concentrate the quantity of peptides in defined theoretical forms. However, applied with denaturation and/or protease-deficient trypsin, Lys-C, Glu-C, Glu-N, and Asp-N, aiming to concentrate the quantity of peptides in defined theoretical forms.

In-solution digestion requires low protease-to-substrate ratio (1:50 or lower, m/m) to reduce contamination, thereby operates near-entirely in the enzyme-deficient, diffusion-limited, low-efficiency region (7); slow reaction rates of compromised trypsin in urea amplify such deviation from ideal. In-solution trypsin (2 M urea) digestion has also been extended by combining with other enzymes such as Lys-C (8 M urea) (17), pepsin (18), Asp-N (19), and PNGase F (pre-digestion (19)). Increasing evidence suggests multispecificity, often via using multiple proteases, is necessary for reliable PSM quantitation of lysates (13, 20). Pressure cycling technology speeded tandem urea-solution Lys-C/trypsin digestion of tissue lysate into 6 h, and strict cycle control enhanced peptide reproducibility (21).

Alternatives to Traditional In-solution Digestion—

On-filter In-solution Digestions (FASP and in-StageTip)—

Avoiding complete precipitation/resolubilization, the filter-based in-solution digestion method FASP (22, 23) achieved landmark improvement in digestion yield (up to 50%) by keeping membrane proteins in solutions during SDS removal. Following SDS protein solubilization, FASP applied dialysis (dial)-filtration with small molecular-weight-cut-off filters to gradually replace SDS in solution with high urea, performed in-solution digestion(s) overnight on filter, then separated peptides from enzyme and undigested proteins via centrifugation. FASP and its derivatives that added deoxycholic acid (2014 eFASP (24)) or further extended the on-filter reactor for reactions such as deglycosylation (2010 N-glyco-FASP (25)), have been widely applied since 2009. However, ~50% protein sample resists digestion likely because of inevitable re-aggregation upon detergent depletion, and although overall protein identification is improved, peptide reproducibility remains low (23). The laborious manual operation is prone to contamination and sample loss. Abandoning SDS, the latest version of FASP, 2014 in-StageTip (26), used smaller tips sealed with C18 disk that doubles as micro-filter besides desalting, and combined protein deoxycholate solubilization, reduction, alkylation, and overnight in-solution digestion in these tips in 96-well format. However, this method still mandates manual processing, hours of digestion, and centrifugation (which interrupt automation streamline), and inherits all problems intrinsic to the denaturation-trypsin in-solution paradigm; ionic and high in cmc (critical micelle concentration; 1% m/v, 4.2× cmc was 25 mN), sodium deoxycholate incurred an extra step of post-digestion removal with ethyl acetate extraction and centrifugation (26), further obstructing automation.

Protein On-bead Digestions—Fixing membrane proteins on beads following specific affinity-enrichment (27, 28) effectively alleviated interprotein aggregation, such as during detergent removal (27), and greatly improved digestion and identification of membrane proteins using Lys-C and trypsin urea solutions (27). However, information of unbound proteins is lost. In the 2014 Single-Pot Solid-Phase-Enhanced Sample Preparation (SP3) method, hydrophilic magnetic beads adsorbed proteins and peptides less discriminatingly, and simplified solution separation for pre-digestion SDS removal and post-digestion peptide cleanup (29). However, common to both protein on-bead strategies, inevitable intraprotein aggregation may limit digestion access, and peptides remain mixed with protease solution.

Immobilized Enzyme Reactors for Fast Reproducible Digestion—Numerous micro-reactors of enzymes, including trypsin (30–32), trypsin/Lys-C (33), pepsin (34, 35), and PNGase F (36–38), adsorbed or covalent-bonded on various solid supports, have been described for peptides and/or glycans studies, mainly for purposes of faster reaction and easy product separation. They demonstrated advantages over in-solution methods, but mostly using simple soluble proteins in high denaturants such as urea, guanidine, or organic solvent.

Covalently immobilized pepsin columns emerged in 2002 for HDX to rapidly digest proteins at low pH low temperature to minimize D-label back-exchange, without complicating spectra with protease peaks (34). With high urea or guanidine denaturant (1.5–3 M, some studies also added 500 mM TCEP) at pH 2.5, pepsin column successfully digested most soluble proteins tested at high (over 90%) coverage and high peptide reproducibility within seconds-minutes, and is widely adopted in HDX (34, 35, 39, 40). HPLC-grade pepsin column’s advantages are multifold and attributable to its operation as a high-surface-concentration plug-flow reactor, in retrospect (7). First, immobilizing pepsin on POROS beads can increase effective surface concentration to over-1-mm scale (34), near-thousand-fold higher than the 2 μm (or 0.1 mg/ml, 50 kDa)
protein typically applied, thereby allowing pepsin column to operate well within the abundance-unbiased high-efficiency region, unrestricted by total sample size (7). Second, flow drive overcomes the problems of product inhibition caused by slow diffusion intrinsic to in-solution incubation, and affords a precise control of the fast product formation from nonspecific protease, which is unconceivable to in-solution and beads-based methods. Third, typical column size, flow rate (1–2 mm i.d. × 20 mm, 25–200 µl/min (6)) and pressure fit with HPLC system for full automation. Trypsin POROS column also showed efficient real-time digestion of soluble proteins online in organic solvent (30–32). However, until 2010 (6), protease columns met no success with membrane proteins, and for years GPCR β2AR coverage remained below 25% (6, 41), preventing useful HDX mapping.

Alternative Detergents for Protein Extraction and Digestion—Other advances were achieved by changing detergents. Traditional detergents for protein extraction in proteomics, such as ionic SDS and poly-disperse Triton X-100, inhibit protease digestion and/or peptide HPLC MS, but are difficult to remove. Alternatively, novel protease-compatible use-and-shred detergents such as RapiGest (42), could increase membrane protein sequence coverage from 10–20% to about 30% in global proteomics (42), and are increasingly applied with success in place of, or after (hours of dial-filtration), SDS or Triton extraction since 2007. However, these detergents vary solubilization power with proteins (some instantly precipitate proteins such as cytochrome c oxidase, CcO), require chemical cleavage post-digestion—difficult for fast full automation, and some cleaved products contain amines that interfere with TMT. A latest acid-cleavable analog of SDS, anionic MaSDeS, showed SDS-like strong solubilization power on tissues, though it is severely denaturing and forms amine products (43, 44). Amphipolymers (amphipols) can support protein extraction and digestion, but most aggregate below pH 3–4 and require removal pre-HPLC by precipitation and centrifugation (45), interrupting automatatable workflow. Several recent studies compared numerous detergents for digestion, but most remained confined to industrial poly-disperse and ionic detergents that deactivated proteins and required removal; other attempts combined harsh detergents, organic solvents, high urea, high pressure, and high temperature (46).

Despite these incremental advances in membrane protein coverage, obvious obstacles endured. First, hydrophobicity and aggregation persisted. Second, PTMs remained mostly unaddressed. Third, all these methods incur laborious manual operation, centrifugation (except for magnetic beads) and hours of processing before HPLC injection, interrupt automation, and remain unable to meet the minutes’ time window allowed by HDX pipeline. Central to these method designs are the common observation and belief that: Detergents are incompatible with the protein-peptide-RP HPLC-ESI MS/MS pipeline, and must be avoided or removed at some point pre-HPLC. Further, the concepts of protein solubilization versus denaturation, and protease digestion reliability versus theoretical specificity, were frankly undistinguished. Consequently, Path 2 (Fig. 1), direct pan-PTM and HDX structural mapping of native functional human membrane proteins—the established high-priority therapeutic targets urgently awaiting precise structure-function roadmaps—remained a formidable field to bottom-up proteomics.

Detergent Selection for Functional Proteomics: The Schrödinger’s Cat Scare—For protein states, we see what the method presents. As a primary tool to retrieve TM proteins from lipid bilayers for solution-based analysis, detergents present proteins in detergent-protein micelles, and inevitably—more or less—affect their function and structure integrity. Rather than presolubilization cross-linking that may disturb the system and complicate analysis, functional proteomics predominantly counts on detergents to preserve proteins’ authentic conformation and non-covalent interacting network, during solubilization and enrichment. Thus ideal detergents ought to: (1) solubilize membrane proteins, (2) maintain proteins’ native structure and interactome, and (3) reduce downstream cleanup that loses samples to re-aggregation and biases against hydrophobic components.

Although current proteomics literature frequently mixed detergents’ solubilization strength with denaturation severity (47), they are in fact two distinct concepts—a premise for functional proteomics. Detergents vary vastly in their effects on protein activity despite solubilization (Table II), ranging from severe (SDS), partial (CHAPS), mild (Triton X-100), to minimal denaturation. Typically 2–5× cmc is required to keep proteins soluble, and more in extraction because of more proteins and lipids in membranes. At such high amounts, ionic (including zwitterionic) and poly-disperse detergents overwhelm peptide reversed-phase (RP) HPLC and ESI, via forming positive fixed charges or strong adducts at acidic pH. Current norm believes detergents must be removed before digestion or HPLC (4, 5, 22), at unavoidable sample loss.

Rigorous membrane protein functional interactome studies typically compared multiple detergents—as they produced vastly variable, partially overlapping protein populations—and emphasized proteins that overlapped or were unique (48, 49). Subsequent analyses resorted to detergent removal by over-night RapiGest exchange-cleavage (48) or by protein precipitation-resolubilization (49), and digestion with trypsin solution (48) or beads (49).

Broadly defined as “a reagent to wipe away” and chemically also termed surface-active agent (surfactant), detergent is characterized with being amphiphilic, and the hydrophilic domain can assume various shapes: one pole of a rod (SDS, Triton X-100), one face of a bulky molecule (CHAPS, deoxycholate), or one side of a linear molecule (amphipols) (Table II).

High-purity n-dodecyl-β-D-maltopyranoside (DDM) was synthesized and first shown to be unusually effective for both solubilizing and stabilizing active membrane proteins—attributable to its micelle size and rigid tight-packing nonionic head.
group—by the Ferguson–Miller group in the 1980s (50–53), using animal tissues then cell cultures. Subsequently, DDM-based extraction and/or purification have critically enabled high-resolution crystallization of most of the over-500 diverse membrane proteins in the Protein Data Bank, including GPCRs, multisubunit ion channels, neurotransmitter transporters, and respiratory complexes (54–60); DDM’s recent “X”-shaped dimeric derivative, maltose-neopentyl glycols (MNG) type 3 (MNG-3, lauryl MNG), stabilized hGPCR-Gs complex for crystallization (55, 61). For mammalian signaling TM proteins native to 20–25%-cholesterol lipid bilayer habitats (by weight and by count) (62–64), the Stevens group found adding cholesteryl hemisuccinate (CHS) to DDM forms wider bicelle-like DDM/CHS micelles that crucially stabilize human GPCRs for activity and crystallization (54, 65). DDM/CHS also proved effective to extract and purify active synaptic hetero-pentamer hGABAAR from HEK293 (66); seconds of DDM/CHS sonication afforded over 90% solubilization of HEK293 membranes (66) and no visible pellet post-centrifuge for rat brains (personal communication, Feb 2015). Decyl MNG/CHS stabilized hLGIC (67). By contrast, traditional ionic or poly-disperse detergents, such as SDS, CHAPS, cholate, deoxycholate, C12E9, Triton X-100, Tween 20, and Brij 35, as well as nonionic DM, OG, and digitonin, failed to emulate despite decades of pursuit (51–53, 68). Ionic CHAPS and cholate consistently deacti-
vated hGABA<sub>N</sub>R (66) and mammalianCcO (52) by several-fold compared with DDM and DDM/CHS. Latest nonionic detergents that combined two to three DDM’s head groups or three glucosides with a steroid body markedly stabilized proteins for crystallization (68). By replacing other detergents, various amphipols and beta-strand peptides also emerged effective to chaperone proteins for high-resolution EM (69, 70).

However, these detergent-driven advances in membrane protein technology since 1980 circulated largely among the crystallography community, not proteomics. Inheriting industrial-grade detergents such as SDS and Triton X-100, historically applied for cheap price, proteomics has suffered huge penalties in quality and cost. Using ESI MS experiments of intact model protein solutions (infusion and flow injection), a pioneering 1994 study reported several nonionic, zwitterionic, and anionic detergents tolerable by ESI MS, in acetic acid/acetonitrile/H<sub>2</sub>O spray (top 7: n-dodecyl glucoside, n-hexyl glucoside, CHAPS, cholate, and three equally scored DDM, OG, and octyl thioglucoside), and in H<sub>2</sub>O spray (top 5: n-dodecyl glucoside, n-hexyl glucoside, OG, n-octyl sucrose, and n-dodecyl sucrose) (71).

We further considered detergent compatibility with protein extraction, protein activity, protease activity, RP-HPLC, data-dependent MS/MS competition and peptide labeling. We propose that beyond tolerance by intact protein MS, detergents may be cultivated as a tool for both membrane protein extraction/purification and digestion modules in proteomic pipeline. Detergents as a tool for membrane protein extraction and purification (discussed above) emphasize maintaining protein native conformations and activity at high yield (often with additives), and providing near-equal grounds of protein states to correlate with other technologies such as crystallography, EM, and function assays (Fig. 1). Detergents as a tool for digestion put emphasis on maintaining protease activity, solubilizing and stabilizing substrates against inter- and intraprotein aggregation over a wide pH range, and minimizing interference with labels (such as HDX, oxidation and TMT), peptide RP HPLC, or ESI MS/MS, but less on native conformations, as digestion often desires acidic pH and proteins are dissected rapidly. Detergents as a tool for united proteomic pipelines emphasize optimal compatibility with all modules.

Based on detergent properties—including charge state at acidic pH, mono- or poly-dispersion, and cmc (count ratio over protein at 5× cmc: 1 μM protein)—and their documented performances in membrane protein extraction, activity maintenance, and high-resolution structures, particularly experiences gained from CcO, hGPCR, and hLGIC complexes, Table II analyzed both common and recent detergents for their utility/compatibility potentials to various steps in integrated functional membrane proteomic pipelines.

**Direct Flow/DDM-facilitated Digestions for HDX, Deep Sequencing, and Quantitation—**

**Discovery of Omnicompatible Detergents as a Pivotal Tool to Fully Automate Deep and Direct Membrane Proteomics—**

Breakthroughs in membrane protein digestion started in 2010 for HDX dynamic structure studies, arguably the ultimate digestion challenge, by including detergents that complied with structure/protease/HPLC MS (6), or by applying membrane scaffold proteins (MSP) that kept proteins in nanodiscs followed by cholate (which entailed high-pressure ultra-performance liquid chromatography, UPLC) (72, 73), to solve aggregation. Thinking outside the box of denaturation and detergent removal, we discovered that nonionic low-cmc amine-free DDM—the most protein nature-promoting and crystallization-successful detergent to date—was completely compatible with structure/protease/TMT/HPLC MS/MS, and devised DDM-based protein extraction, purification and digestion methods that cleared all these challenges of membrane proteins (6, 7, 66). DDM-low-TCEP (DLT, for HDX) and flow/detergent-facilitated protease and de-PTM digestions (FDD, for deep sequencing and quantitation) both completed peptide preparations within seconds-minutes, at robust coverage and peptide reproducibility supporting full automation, demonstrated by using hGPCR (6) and hLGIC complex (7), respectively.

Advantages of DDM as a tool for membrane proteomics are multifold and pivotal to achieving fully automated deep and direct proteomic analyses of membrane proteins in physiological states. First, for protein extraction/purification, DDM combined with CHS to mimic cholesterol in human cells, can extract membrane proteins at high activity, and high efficiency when applied with co-micelle-promoting slow addition or sonication (66). DDM and its derivatives that critically enabled the purification and crystallization of myriad active membrane proteins (60, 51, 54–57, 61, 74, 75), present proteins in functional states in solution, thus providing a common ground for proteomics to bridge functions directly with protein PTMs, dynamic conformations, and static structures resolved by crystallography (Fig. 1). Second, as a tool for digestion, DDM preserves—not destroys—the activity and solubility of both enzymes and protein substrates, thus facilitates effective catalytic contact and unbiased flow of hydrophobic species during digestion, and obviates problematic re-aggregation, high urea, subsequent desalting, and reconcentrating. DDM is effective over a wide pH range, thus can broadly simplify diverse digestion formats such as: enzyme column, protein on-bead, in-solution, and in-nanodisc, by replacing conventional detergents, urea or guanidine. Third, for overall pipelines, amine-free nonionic DDM supports downstream HPLC, MS/MS competition and TMT reactions, can be included throughout the cell lysis-protein extraction-(HDX)-digestion-(TMT)-HPLC-MS/MS process without removal, thus confers the workflows with no interruption, no sample leakage and high sensitivity. DDM at 0.4–1 mM (0.02–0.05%, 2.3–5.7× cmc) over 1 μM proteins (0.04–0.28 mg/ml) was well tolerated (6, 7).

At even lower cmc (~9, ~10, and 36 μM) (61), nonionic MNG-2, MNG-3, and decyl MNG are expected compatible
with RP HPLC MS as well; likewise are the low-cmc nonionic facial amphiphiles (Fig. 2B and Table II). MNG-3 greatly stabilized protein complex in solution for crystallization (61), we found its membrane solubilization power for HEK293 was comparable to DDM basal (66), thus MNG can serve as a valuable control to cross-check DDM extraction results and as another tool for digestion.

**DDM-low-TCEP Digestion for HDX**—Instead of conventional 3 M urea, 1.5 M guanidine, or 500 mM TCEP, the 2010 DLT method developed a digestion solution of 0.02% (m/v) DDM, low 15 mM TCEP and 100 mM NaH₂PO₄·HCl pH 2.4 (final 7.5 mM TCEP), to quench membrane protein HDX (DDM/CHS/glycerol protein buffer in H₂O or D₂O) for direct digestion, online in a fully automated workflow (6). DLT with pepsin column achieved the first comprehensive HDX profile of 7TM hGPCR at near 90% coverage, using a PTM-deprived minimalist β₂AR developed for crystallization (6). These peptides were each reproduced by several hundred independent digestions of apo and various ligand-bound β₂AR from multiple batches of purifications (Ref (6) and unpublished data). The DDM solution also successfully reproduced high coverage for other GPCRs in other laboratories (76).
However, devised for direct HDX, DLT did not address extensive PTMs and Cys, as they did not pose a problem in the \(\beta_2\)AR crystallizable construct: All PTM-bearing domains were either genetically truncated (intracellular C-terminal tail and loop 3) or deglycosylated (two adjacent extracellular-domain N-glycosylations) in purification (6). Although 7.5 mM TCEP allowed detecting peptides bearing free Cys, without alkylation, they were the weakest, and the C–C linked versions were detected at higher ion counts (6). Also designed with a desalting trap and high-flow-rate HPLC and ESI (50 l/min), DLT-HDX’s peptide signals were close to baseline in HPLC-ion count chromatograms, and did not appeal to broader applications (Fig. 3). 

**Flow/DDM-facilitated Protease and De-PTM Digestions for Deep Sequencing and Quantitation**—Because DDM obviated high salt, by removing the desalting trap and adopting capillary HPLC column and nanoESI, a no-trap-nanoESI upgrade of DLT remarkably increased peptide ion sensitivity (66). When applied to full-length hGABA\(_A\)R hallmarked with Cys-loops, it produced up to 80–84% coverage with a 75-min HPLC gradient (66), but gaps of over 15% persisted at Cys and potential PTM sites.

Addressing PTMs and Cys, the 2015 FDD method further combined immobilized pepsin column, PNGase F column, Cys-selective reduction and alkylation, and greatly improved peptide metrics. A novel integrated digestion paradigm, FDD’s components each targeted the key challenges for human signaling TM proteins—hydrophobicity and intricate PTMs—inaccessible for decades by conventional methods. (1) FDD’s tandem flow reactor format achieved fast, effective, reproducible, and controllable reactions for both proteolysis and de-PTM; (2) DDM resolved aggregation caused by TM hydrophobicity and promoted unbiased flow through all reactors; (3) de-PTM PNGase F column solved thorough mapping of N-glycosylation sites that turned out to crosstalk with other PTMs and critical for functions. At minutes of total preparation time, FDD achieved up to 99% coverage in one run, high peptide reproducibility and low artifacts for all subunits of hGABA\(_A\)R/HEK293 (7) (Fig. 4A and Fig. 3C). Importantly, the FDD-no-trap-nESI platform increased peptide ion sensitivity by four orders of magnitude from 2010 DLT-trap-ESI (Fig. 3), and discovered DDM is no longer a compromise, but a superior tool broadly applicable to nESI-based proteomics. FDD further demonstrated direct TMT labeling and analysis (7) can be integrated in a fully automatable membrane proteomic workflow (Table I and Fig. 2).

FDD presents multifold advantages for bottom-up membrane proteomics, and critically supports the next-level quantitative-structural integration. First, FDD transformed the days-long aggregation-plagued paradigm into immediate digestion and measurement of active protein solutions—in seconds–minutes total peptide preparation time—at unprecedented coverage, reproducibility, and authenticity. This allowed both the minutes-windowed HDX and the deep PTM/interactome mapping to look at the near-identical states of proteins and peptides, thus broke barriers against integrating quantitative-structural membrane proteomics. Second, FDD eliminated precipitation, filtration and centrifugation central to the denaturation-, tip- or filter-based in-solution traditions. This enabled a robust, instant, versatile protein-(HDX)-peptide-(TMT)-HPLC MS streamline for uninterrupted automation. Together, FDD made Path 2 immediately achievable (Fig. 1) and may improve Path 1 in speed, depth, accuracy, sensitivity, and throughput (Table I).

Striking peptide metric increases in DLT and FDD are not artifacts of database search, but reflect real changes in peptide production. First, digestion methods—urea-based trypsin or pepsin versus DLT-pepsin digestion of GPCR (6), and DDM-facilitated pepsin solution, beads versus flow reactor format (66)—were compared on the equal grounds of target

---

**FIG. 3.** At much higher peptide ion sensitivity, the 2015 FDD-no-trap-nESI method proved DDM can be a superior tool for broad nESI-based membrane proteomics. Representative HPLC MS traces of (A) 2010 DLT-HDX-trap-ESI of hGPCR apo \(\beta_2\)AR, typical to the data for (6), B, FDD-no-trap-nESI of hGABA\(_A\)R, and (C) independent digests of H/D versions of hGABA\(_A\)R using FDD-no-trap-nESI’s HDX module showing good reproducibility. All spectra were acquired with orbitrap analyzers in Thermo Exactro (A), Q-Exactive (B), or LTQ-orbitrap XL (C) mass spectrometer. Experiments used similar concentrations of protein and detergent, but A loaded several-fold more total sample to HPLC than B and C. Figure was adapted from (7).
sequence search and parallel experiments. Second, spectra search against human proteome maintained the high peptide number and coverage of target proteins (7) (Fig. 4B). Third, identified proteins showed average sequence coverage higher than current 25% benchmark (8, 42) (Fig. 4B). Further, compared one-to-one with 7TM hGPCR and hGABAAR, most proteins in cell lysates are technically easier, thus unlikely to pose further obstacles.

Concrete reach to PTMs—including N-glycosylation on Cys-loop and endogenous M-oxidation—proved crucial for discovering function clues inaccessible to methods with sporadic coverage and severe artifacts (7). Extensive PTMs in human signaling TM proteins, such as high-occupancy N-glycosylation, had remained largely a target of observation, but rarely exploited as a means to aid overall peptide formation and detection, partly for fear that PTM enzymes overwhelm real samples (though PNGase F was used to release enriched glycopeptides (77)). Predigestion PNGase F solution incubation indeed identified more proteins in yeast cell wall (19). But emerging complicated N-glycosylation scenarios in mammalian TM proteins suggest, treating peptides post-digestion is more likely efficient and complete than treating proteins (7, 66).

**Extended Applications of FDD Format**—As a general platform to prepare membrane protein peptides, FDD's elements—enzyme flow reactors, omnicompatible detergents, and de-PTM—are flexible to adapt to nearly all levels of proteomics and support myriad enzymes and new detergents (Fig. 2C). Both the DDM/CHS-extraction and FDD digestion strategies are compatible with upstream enrichment of cell fractions, proteins, protein nanodiscs, downstream peptide enrichment (DDM supports IMAC (6, 54, 57, 74, 75)), fractionation, TMT, and HPLC. Excitingly, besides
DDM/CHS, MNGs and several recent detergents superb for crystallography and EM likely fit FDD as well (Table II), suggesting the proposed proteomics-EM-crystallography-function integration is not an exception, but an increasingly achievable new paradigm for membrane proteomics (Fig. 1). This flexibility also allows matching digestion detergent with the one used in protein preparation to avoid multiple HPLC peaks.

**Membrane Protein Digestion: Protease Reliability beyond Theoretical Specificity**—Confounded by membrane proteins’ unique hydrophobicity and tendency to aggregate, protease’s empirical reliability has proved distinct from its theoretical specificity over primary sequence. How to apply the protease appears to matter more than theoretical specificity for achieving reliable digestion. Several recent digestion advances can be interpreted via improving substrate–protease contacts, such as the protease-mild substrate-solubilizing RapiGest (42), and the pressure cycling technology to improve reaction mixing in urea (21), though all under the detergent-removal/avoiding paradigm. For this purpose, an HPLC-grade flow-propelled high-concentration immobilized enzyme micro-column reactor is arguably the most effective. Omnicompatible detergents provided the bridge for membrane proteins to take advantage of this format, to reach complete reproducible digestion and seconds–minutes scale under ambient physiological solutions, temperature, and pressure, obviating any brutal force.

For specific proteases, effective substrate-protease contacts shall help overcome the deviation from theoretical specificity, and finally achieve the “peptide concentrating” strategy described above. DDM-facilitated trypsin beads already showed complete digestion of a multihelix protein within minutes (supplemental Fig. S1A) contrasting overnight in urea solution, thus DDM-facilitated trypsin column is expected effective with membrane proteins too (Fig. 2C). Applying multiple specific proteases in the format of HPLC-grade tandem columns or mixed-beads column (minutes) shall be more efficient than combining days of individual urea-solution digestions (Fig. 2C).

Challenging conventional belief, fast multi-specific protease proved able to be highly reproducible and reliable—when applied in the DDM-facilitated HPLC-grade enzyme column format (6, 7). Reproducible peptide ladders provide desired redundancy and bypass specific proteases’ cleavage gaps. Pepsin column rapidly generated comprehensive coverage including TM domains, with redundant overlapping peptides four to 20 residues long, useful for HDX and PTM mapping (6, 7). Nepenthesin-1, a recent pepsin alternative, has different multispecificities, max activity at pH 2.5, low autolysis, low interference with trypsin cleavage impeding peptide-centric quantitation, (3) when proteins such as myoglobin resist trypsin digestion (supplemental Fig. S1B Lane 3), and (4) when acidic digestion is desired to minimize nonenzymatic deamination (82) and scrambling of native PTMs such as C-C bonds (83). The large number of peptides is within capacity of matured HPLC separation.

**CONCLUSIONS**

Shedding the decade-long cliché of denaturation, detergent removal/avoidance and laborious processing, the detergent- and reactor-based digestion strategies demystified the technical challenges of membrane proteins, shifted the day(s)-long paradigm into seconds–minutes scale, and cleared barriers against real-time integration in proteomic pipelines. These digestion methods empower membrane proteomics to grow beyond protein identification to the next level: integrative structure-function mechanisms of membrane proteins, via direct quantitative PTM and structural mapping in their physiological states, at unprecedented coverage, throughput, sensitivity, and accuracy (Fig. 1). At the surge of ambition and efforts to tackle membrane protein-central human proteome, signaling interactome, and dynamic structural mechanisms of TM receptors for better therapeutics, they afford a timely tool box of solutions to accelerate bottom-up proteomics to achieve these goals.

**Acknowledgments**—I thank Dr. Shelagh M. Ferguson-Miller for discussions on detergents and this manuscript. I thank Dr. Keith W. Miller for providing a CYP11B1 sample used in supplemental figure.

*This work was supported in part by the National Institute of General Medical Sciences (GM 58448, K.W.M.).

[1] This article contains supplemental Fig. S1.

[1] To whom correspondence should be addressed: Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, 32 Fruit Street, Boston, MA 02114. Tel.: (517) 944-0842; Fax: (617) 724-8644; E-mail: xi.zhang.edu@gmail.com.

**REFERENCES**

Real-time Membrane Protein Digestion for Future Proteomics