Carnivorous Nutrition in Pitcher Plants (*Nepenthes* spp.) via an Unusual Complement of Endogenous Enzymes

Linda Lee,† Ye Zhang,‡ Brittany Ozar,† Christoph W. Sensen,‡ and David C. Schriemer*†

†Department of Biochemistry and Molecular Biology and the Southern Alberta Cancer Research Institute, University of Calgary, Calgary, Alberta T2N 4N1, Canada
‡Institute of Molecular Biotechnology, Graz University of Technology, Graz 8010, Austria

Supporting Information

ABSTRACT: Plants belonging to the genus *Nepenthes* are carnivorous, using specialized pitfall traps called “pitchers” that attract, capture, and digest insects as a primary source of nutrients. We have used RNA sequencing to generate a cDNA library from the *Nepenthes* pitchers and applied it to mass spectrometry-based identification of the enzymes secreted into the pitcher fluid using a nonspecific digestion strategy superior to trypsin in this application. This first complete catalog of the pitcher fluid subproteome includes enzymes across a variety of functional classes. The most abundant proteins present in the secreted fluid are proteases, nucleases, peroxidases, chitinases, a phosphatase, and a glucanase. Nitrogen recovery involves a particularly rich complement of proteases. In addition to the two expected aspartic proteases, we discovered three novel nepenthensins, two prolyl endopeptidases that we name neprosins, and a putative serine carboxypeptidase. Additional proteins identified are relevant to pathogen-defense and secretion mechanisms. The full complement of acid-stable enzymes discovered in this study suggests that carnivory in the genus *Nepenthes* can be sustained by plant-based mechanisms alone and does not absolutely require bacterial symbiosis.

KEYWORDS: *Nepenthes*, transcriptomics, carnivory, mass spectrometry, fluid, enzymes

INTRODUCTION

Carnivorous plants use specialized trapping structures to attract, capture, digest, and absorb nutrients from insect prey and other sources of nitrogen, phosphorus, and minerals.1−5 Trapping mechanisms include snap traps (e.g., *Dionaea*, also known as Venus flytrap), flypaper or adhesive traps (*Drosera*, *Pinguicula*), sucking bladder traps (e.g., *Utricularia*), and pitfall traps (e.g., *Nepenthes*). Plants belonging to the genus *Nepenthes*, also known as pitcher plants or monkey cups, are particularly intriguing. The genus consists of over 100 species found mostly in Southeast Asia and other tropical regions.6 The *Nepenthes* pitcher buds from the end of the leaf and progressively forms a pitfall trap consisting of three sections: a slippery upper rim called the peristome that is involved in attracting and trapping prey; a slippery, waxy inner wall to trap and prevent escape; and a bottom pit filled with an acidic viscoelastic fluid used to digest the trapped prey. Several structural and chemical variations of the trap seem to be adapted for a variety of noninvertebrate prey, including plant detritus and even tree-shrew feces.7−9 The traps contribute a remarkably high percentage of the total nutrient requirement for the plant10 and raise interesting questions regarding the processes involved, particularly those related to prey digestion.

A common feature of all traps is the containment and presentation of the digestive fluid. Glands line the bottom section of the pitcher and perform numerous functions, including the detection of stimuli, absorption of nutrients, and secretion of digestive enzymes and acid into the pitcher fluid. Attempts in recent years to identify and characterize the protein composition of the *Nepenthes* pitcher fluid yielded a surprisingly low number of enzymes. On the basis of this body of work, it appears that the most abundant proteins in the *Nepenthes* digestive fluid are Nepenthensins 1 and 2 (Nep1/2).8,9 These are noncanonical aspartic proteases that exhibit broad cleavage specificity, high thermal stability, and optimal activities at pH as low as 2.5.9−12 Aside from these two aspartic proteases, a handful of other proteins have been identified in pitcher fluid, including a β-1,3-glucanase, a β-D-xylosidase, chitinases, a peroxidase, pathogenesis-related protein PR-1, a thaumatin-like protein, and a ribonuclease.13 Even this minimal composition begins to suggest a concerted mechanism for nutrient acquisition and inhibition of microbial pathogens.13−20

Prior to modern proteomics methods, other enzymatic activities were either detected or suspected. Activities arising from cysteine proteases, phosphatases, esterases, oxidoreductases, and lipases have been attributed to the *Nepenthes* fluid.8,19−25 Nevertheless, the identity and source of most of these enzymes are not known. The *Nepenthes* fluid harbors a diverse microbial community, some of which may contribute enzymatic capacity to the fluid and form a symbiotic relationship with the plant, although their relative contribution...
to prey digestion is still under investigation. However, studies have questioned whether the bacterial load in the trap has a role to play in the processing of insect prey, prompting a closer inspection of fluid composition.

Although proteomics methods have been applied to the fluid, the complement of proteins should not be considered complete. A thorough identification of all the proteins present in the *Nepenthes* digestive fluid is challenged by two major problems: the unusual amino acid composition of the proteins and the limited representation of carnivorous plants in the genomic/protein databases. Many of the proteins identified in the *Nepenthes* digestive fluid thus far have a low frequency of Lys/Arg residues. These characteristics suggest that standard genomic/protein databases. Many of the proteins identified in the *Nepenthes* digestive fluid thus far have a low frequency of Lys/Arg residues. These characteristics suggest that standard

Moreover, the identification of the proteins is complicated by the lack of a complete sequence of the *Nepenthes* genome and transcriptome. Although some transcriptomics resources are beginning to emerge, and used in a preliminary tryptic-based proteomics workflow, an alternative strategy is required to test for the completeness of the compositional analysis. To acquire comprehensive insight into the enzymatic components secreted by the plant, we sequenced the *N. rafflesiana* transcriptome and coupled it to proteomic analysis of the fluid using a combination of approaches. In addition to trypsin, we took advantage of the robust proteolytic capacity of the fluid itself to increase the depth of coverage. Using these complementary approaches, we were able to confirm many of the known fluid proteins. We were also able to identify novel proteins that likely function in prey digestion and defense.

### EXPERIMENTAL METHODS

#### RNA Extraction

*N. rafflesiana* giant plants were purchased from Urban Bog (Canada) and grown in a small terrarium in a 15:9 h light:dark photoperiod. The plants were fed with 1 or 2 *Drosophila* sp. per pitcher the day before the pitchers were to be harvested for RNA extraction. Prior to harvesting the pitcher for RNA extraction, the digestive fluid was decanted, and the pitcher was was deionized water to remove partially digested material and other debris. The bottom one-third of the pitcher containing the secretory glands was excised, frozen in liquid nitrogen, and ground to a fine powder in the presence of liquid nitrogen. The total RNA was extracted using a modified CTAB protocol described by Meisel et al. In brief, the ground pitcher was lysed in warm CTAB extraction buffer (2% w/v cetrimonium bromide, 100 mM Tris-HCl pH 8, 20 mM EDTA, 1.4 M NaCl, 1% w/v polyvinylpyrrolidone MW 40,000, 0.05% w/v spermidine trihydrochloride, and β-mercaptoethanol) and extracted several times with chloroform-isooamyl alcohol (24:1). The RNA was then precipitated from the aqueous phase with 2.5 M LiCl, followed by chloroform extraction and precipitation in ethanol. The quality and integrity of the RNA was assessed on a 2200 TapeStation (Agilent) to have a RIN score >8.

#### Transcriptome Sequencing

Illumina RNA sequencing of *N. rafflesiana* transcriptomes was performed by Omega Bioservices (Omega Biotek, Inc., Norcross, GA). RNA samples were processed using the Illumina TrueSeq Stranded Total RNA Sample Prep kit (Illumina, San Diego, CA) following the manufacturer\'s protocol. Briefly, 1 μg of total RNA was treated with Ribo-
was released after 1 spectrum and 0.2 min to avoid reacquiring MS/MS data for the same precursor. The data were searched against the 6-frame translation of our *N. rafflesiana* transcriptome databases using Mascot. The Mascot search parameters for the tryptic digested samples were a maximum of one missed cleavage, carbamidomethyl (C) as a fixed modification, methionine oxidation as a variable modification, a 20 ppm peptide mass tolerance, and a 0.2 Da fragment mass tolerance. Output was filtered for peptide hits with *p* < 0.05 (ion score \( \geq 24 \)).

**Gel-Free Fluid Proteomic Analyses**

For the proteome of the *N. ventrata* pitcher fluid to be analyzed, the fluid was digested either with trypsin (specific digest) or *N. ventrata* fluid (nonspecific digest) using a modified FASP protocol. Briefly, 15 \( \mu \)g of 10\( \times \) concentrated *N. ventrata* digestive fluid was reduced with 100 mM DTT followed by alkylation with 50 mM iodoacetamide under denaturing condition (8 M Urea, 100 mM Tris-HCl pH 8) on a YM-10 microcon (Millipore). For digestion with trypsin, the reduced/alkylated fluid was buffer exchanged to 50 mM ammonium bicarbonate pH 8 and digested with trypsin (Promega; enzyme to substrate ratio of 1:100) at 37 °C overnight followed by quenching with 0.1% trifluoroacetic acid. For nonspecific digestion, the reduced/alkylated fluid was washed with 100 mM Gly-HCl pH 2.5 and digested with *N. ventrata* fluid (1:30 enzyme:substrate) at room temperature for 40 min. The digests were desalted on a C18 ZipTip (Millipore) prior to analysis by data-dependent LC−MS/MS on an Orbitrap Velos ETD (Thermo Scientific). The samples were analyzed twice using 90 min reversed-phase gradient runs configured for top-10 ion selection using CID in a high/low configuration. One run focused on peptides with a charge state of 1\(^+\), whereas the other run focused on peptides with a charge state of 2\(^+\) and higher. MS scans were acquired over the m/z 350−1800 range with a resolution of 60,000 (m/z 400). The target value was 5.00 \( \times \) 10\(^5\). For MS/MS, ions were fragmented in the ion trap with normalized collision energy of 35%, activation \( q = 0.25 \), activation time of 10 ms, and one microscan. The target value was 1.00 \( \times \) 10\(^3\). Separation was achieved using a self-packed pico-frit column (New Objective, packed with Phenomenex \( 5 \) \( \mu \)m Jupiter C18 beads, 100 mm × 75 \( \mu \)m), using 0.1% formic acid in 3% acetonitrile for mobile phase A and 0.1% formic acid in 97% acetonitrile for mobile phase B (flow rate of 300 nL/min and a 0.5% B/min linear gradient to 45% B). The data were searched against the 6-frame translation of our *N. rafflesiana* transcriptome databases using Mascot. The search parameters for the tryptic digested samples were one missed cleavage allowed, carbamidomethyl (C) as a fixed modification, methionine oxidation (M) as a variable modification, a 10 ppm peptide mass tolerance, and a 0.6 Da fragment mass tolerance, and the output was filtered with a false discovery rate of 1% (peptide hits with \( p < 0.00045 \); ion score \( \geq 42 \)). The search parameters for the fluid-digested samples were similar to those for the tryptic sample, except the search was configured for “no enzyme” specificity and the output set was filtered with a false discovery rate of 1% (peptide hits with \( p < 0.009 \); ion score \( \geq 49 \)). In cases where shorter partial ORF were contained within or overlap with another transcript, the longer assembled transcript was used for characterization.

### RESULTS

**Transcriptome of the Stimulated *Nepenthes* Pitcher**

To support the proteomics-based identification of the protein composition of *Nepenthes* digestive fluid, we performed Illumina RNA sequencing to generate a comprehensive database of the *Nepenthes* transcriptome (Figure 1). We extracted the RNA from the bottom third of a *N. rafflesiana* giant pitcher, where the secretory glands are located, which had been stimulated by feeding with fruit flies to reduce fluid pH and promote the digestive state. Because of the open nature of the *Nepenthes* pitcher, bacteria and fungi as well as fruit flies could be present in the pitcher. However, read pairs corresponding to these contaminating genomes accounted for less than 1% of the total read set (Figure S1 and Table S1) with proteobacteria being the most dominant phylum (Figure 2). After the raw data were cleaned, including the removal of the contaminating genomes, the transcriptome was assembled from 154 million paired-end reads of 100 bp length. A total of 1.36 million contigs >200 bp were assembled with an average length of 1292 bp. The number of total contigs was significantly reduced to 335 K with an average length of 1329 bp (Table S1). When the assembled transcriptome was filtered for contigs >300 bp, the number of total contigs significantly reduced to 335 K with an average length of 1329 bp. The high number of contigs, especially those <300 bp, may arise from small, noncoding RNA and/or errors in assembly. Nevertheless, we maintained the cutoff at 200 bp for this stage of analysis. Searching our transcriptome with tRNAscan did not find any significant tRNA hits, and alignment against the *Arabidopsis* intronic-itergenic database yielded minimal results, indicating that our transcriptome was enriched for mRNA as expected. When we included expression as a criterion, we reduced the number of contigs to 322,208 contigs (>200 bp set). Although only 31% of the expressed contigs could be annotated using the NCBI nucleotide database, these accounted for nearly half of the total number of reads (Table S3). The unannotated contigs accounted for only ~12% of total read set, suggesting that these may be present at very low levels.
(and/or arise from error in assembly). Taxonomical annotation of the contigs using MEGAN revealed that well over half of the transcripts were plant-based (Streptophyta) with some minor contamination from bacteria, fungi, and fruitflies (Figure 2 and Table S2), indicating that our method enriched for Nepenthes transcripts. To enrich for long transcripts and enable an effective proteomics experiment, we filtered the translated ORFs for polypeptides >150 aa in length, which generated 66,459 ORFs for the search.

**Proteome of the Nepenthes Digestive Fluid**

Visualization of the stimulated N. ventrata digestive fluid via silver-stained SDS-PAGE showed the limited complexity of the fluid in keeping with previous work (Figure 3). When searched by long-gradient LC–MS/MS identified 26 proteins in the fluid. Sequence coverage was still low with only 1–4 identified peptides matching to each protein (Table S4); these results are similar to those of a recent study.30 As noted, the low level of coverage could arise from a nonoptimal peptide set due to the low frequency of lysine and arginine residues present in the majority of the proteins identified to date. For example, nepenthesin 2 is relatively abundant in the digestive fluid,9,13,34 but it was not detected in the tryptic digest sample due to the absence of K/R residues in its mature active form. Deglycosylation of the protein sample did not improve coverage (not shown).

To circumvent the restrictions presented with trypsin, we digested denatured N. ventrata fluid with active N. ventrata fluid. The high activity of the fluid, and its value in generating high sequence coverage in hydrogen/deuterium exchange MS experiments, argued for the attempt. At our current level of characterization, the fluid minimally contains aspartic proteases with broad cleavage specificity.11,12,34 Upon LC–MS/MS of the nonspecific digested fluid, we obtained substantially higher coverage with as many as 70 matching peptides for the most abundant proteins. We identified 19 proteins using this approach (Table S5). Nine proteins are common with those identified in the tryptic digestion method; therefore, in total, 36 proteins were identified in the N. ventrata fluid from the combination of tryptic and nonspecific digestion approaches (Table 1). The identified proteins were functionally annotated by homology search against the NCBI nonredundant (nr) protein sequence database or characterized biochemically. For the relative abundance of the identified proteins based on label-free methods, such as emPAI, to be assessed, the tryptic data is particularly unreliable given the paucity of detectable peptides. However, the emPAI values from the nonspecific digests are useful given the much larger peptide set generated. The ranked list in Table S5, therefore, approximates the order of abundance.

Our improved digestion strategy presents a more complete analysis of the fluid proteome when compared to a recent study using only trypsin.30 We identified approximately 25% more...
proteins, some of which are novel. Globally, the majority of the identified pitcher proteins have properties expected of a secreted protein functioning in an acidic fluid. Over 70% of the identified fluid proteins have a pI < 6, consistent with the acidic nature of the digestive fluid (Table 1 and Table S6). The pI is predicted from the deduced unprocessed transcript and may not reflect that of the mature active enzymes. Acid activation of the well-characterized Nep1 and Nep2 involves autolysis at the N-terminus,35 yielding a mature enzyme with a lower pI than the proenzyme. Similar post-translational processing events may also occur for other fluid proteins, possibly resulting in even lower pI values. Consistent with the fact that the fluid proteins are secreted into the pitcher, two-thirds of the identified proteins with an intact N-terminus were predicted by SignalP 4.1 to have a signal peptide.

More specifically, the fluid proteome can be organized into a number of functional classes and subclasses that aid in evaluating the mechanisms of carnivory.

Table 1. Combined List of the Identified Proteins from the Secreted Nepenthes Fluid

<table>
<thead>
<tr>
<th>process</th>
<th>protein</th>
<th>class/function</th>
<th>MW</th>
<th>pI</th>
<th>signal pep.</th>
<th>dig. 1</th>
<th>dig. 2</th>
<th>ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>protein metabolism</td>
<td>Nepenthesin-1, Nep1</td>
<td>aspartic protease</td>
<td>47106</td>
<td>4.7</td>
<td>yes</td>
<td>2</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Nepenthesin-2, Nep2</td>
<td>aspartic protease</td>
<td>46625</td>
<td>4.1</td>
<td>yes</td>
<td>2</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Nepenthesin-3, Nep3</td>
<td>aspartic protease</td>
<td>49092</td>
<td>4.9</td>
<td>yes</td>
<td>3</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nepenthesin-4, Nep4</td>
<td>aspartic protease</td>
<td>48991</td>
<td>4.8</td>
<td>yes</td>
<td>23</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nepenthesin-5, Nep5</td>
<td>aspartic protease</td>
<td>48811</td>
<td>4.6</td>
<td>yes</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neprosin-1, Npr1</td>
<td>prolyl endopeptidase</td>
<td>43161</td>
<td>6.3</td>
<td>yes</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neprosin-2, Npr2</td>
<td>prolyl endopeptidase</td>
<td>&gt;33 kDa</td>
<td>5.2</td>
<td>n/a</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Serine carboxy-peptidase, SCP1</td>
<td>S10 acidic peptide</td>
<td>&gt;47 kDa</td>
<td>4.3</td>
<td>yes</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Purple acid phosphatase, PAP1</td>
<td>metallophosphatase</td>
<td>71268</td>
<td>5.2</td>
<td>yes</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>nucleic acid metabolism</td>
<td>Ribonuclease, S-like, RNaseS</td>
<td>secreted RNase/scavenge phosphates</td>
<td>26224</td>
<td>4.1</td>
<td>yes</td>
<td>3</td>
<td>19, 20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Endonuclease 2, Endo2</td>
<td>S1–P1 nuclease/ cleave RNA and ssDNA</td>
<td>32487</td>
<td>4.9</td>
<td>yes</td>
<td>14</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>polysaccharide metabolism</td>
<td>β-1,3-Glucanase, Glu1</td>
<td>glycoside hydrolase 17/pathogen defense</td>
<td>36752</td>
<td>4.7</td>
<td>yes</td>
<td>6</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Acidic Chitinase, Chit3</td>
<td>glycoside hydrolase 18/class III chitinase/pathogen defense</td>
<td>31086</td>
<td>4.1</td>
<td>yes</td>
<td>12</td>
<td>15</td>
<td>14, 18, 52</td>
</tr>
<tr>
<td></td>
<td>Chitinase, Chit1</td>
<td>glycoside hydrolase 19/class IV chitinase/pathogen defense</td>
<td>29306</td>
<td>4.5</td>
<td>yes</td>
<td>13</td>
<td></td>
<td>13, 53</td>
</tr>
<tr>
<td></td>
<td>DOMON-like domain, Dom1</td>
<td>heme-binding motif/glycoside hydrolases</td>
<td>27995</td>
<td>5.5</td>
<td>yes</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pathogenesis/host defense</td>
<td>Thaumatin-like protein, TLP1</td>
<td>antifungal activity/pathogen defense</td>
<td>23785</td>
<td>5.7</td>
<td>yes</td>
<td>4</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hevein, Hev1</td>
<td>chitin binding/antifungal activity</td>
<td>20752</td>
<td>6.4</td>
<td>yes</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cat. peroxidase 1, Prx1</td>
<td>heme-containing/oxidation</td>
<td>33598</td>
<td>6.4</td>
<td>yes</td>
<td>8</td>
<td></td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Cat. peroxidase 2, Prx2</td>
<td>heme-containing/oxidation</td>
<td>33385</td>
<td>4.6</td>
<td>yes</td>
<td>21</td>
<td></td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Peroxidase, Prx3</td>
<td>heme-containing/oxidation</td>
<td>36354</td>
<td>4.7</td>
<td>yes</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>signal transduction</td>
<td>Leu-rich repeat protein kinase, RP1</td>
<td>See/Th kinase/plant dev./pathogen resistance</td>
<td>66702</td>
<td>5.4</td>
<td>yes</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Leu-rich repeat protein, LRR1</td>
<td>possible protein kinase/ribonuclease inhibitor</td>
<td>69848</td>
<td>4.8</td>
<td>no</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>energy metabolism</td>
<td>ADP/ATP translocase, ADT1</td>
<td>ADP/ATP exchange</td>
<td>&gt;31 kDa</td>
<td>10.0</td>
<td>n/a</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Glyceraldehyde-3-phosphate dehydrogenase, GPD1</td>
<td>NAD-binding/glycolysis/glyconeogenesis</td>
<td>36664</td>
<td>7.1</td>
<td>no</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ATP synthase/H−-ATPase subunit α, ATPA</td>
<td>ATP synthase/membrane-bound proton pump</td>
<td>&gt;18 kDa</td>
<td>5.3</td>
<td>no</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ATP synthase/H−-ATPase subunit β, ATPB</td>
<td>ATP synthase/membrane-bound proton pump</td>
<td>&gt;51 kDa</td>
<td>4.9</td>
<td>no</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>protein transport</td>
<td>Actin, Act1</td>
<td>2-phospho-D-glycerate conversion</td>
<td>&gt;30 kDa</td>
<td>4.9</td>
<td>n/a</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tic20-like protein, Tic20</td>
<td>chloroplast import protein</td>
<td>31449</td>
<td>9.4</td>
<td>no</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Protein transporter Sec1</td>
<td>vesicle trafficking/secretion</td>
<td>34078</td>
<td>9.2</td>
<td>no</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>translation/protein synthesis</td>
<td>Elongation factor 1-alpha, EF1a</td>
<td>GTP-binding elongation factor</td>
<td>49815</td>
<td>9.2</td>
<td>no</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Elongation factor 2, EF2</td>
<td>GTP-dependent ribosomal translocation</td>
<td>95948</td>
<td>6.1</td>
<td>no</td>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>40S Ribosomal protein S14, RS14</td>
<td>protein synthesis</td>
<td>16367</td>
<td>10.6</td>
<td>no</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Caperone protein Hsp90</td>
<td>protein folding</td>
<td>91593</td>
<td>5.0</td>
<td>yes</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>transcription/gene regulation</td>
<td>Histone acetyltransferase, HAC1</td>
<td>acetylate transcriptional regulators, histones</td>
<td>&gt;86 kDa</td>
<td>6.0</td>
<td>no</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Auxin response factor, ARF1</td>
<td>transcription factor/Auxin response factor</td>
<td>&gt;25 kDa</td>
<td>8.5</td>
<td>n/a</td>
<td>19</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

“Trp digestion. Ranking is based on the MASCOT search score. *Non specific digestion using active Nepenthes fluid. Ranking based on MASCOT search score. * Not applicable (n/a) because the protein fragment has a truncated N-terminus.

DOI: 10.1021/acs.jproteome.6b00224
J. Proteome Res. 2016, 15, 3108–3117
Proteases. The most abundant proteins in the Nepenthes fluid are Nep1 and Nep2, which exhibit >90% homology to Nep1 and Nep2, respectively, from other Nepenthes species. Nep1 and Nep2 are noncanonical aspartic proteases with broad cleavage specificity that are stable over a wide temperature range.9−12 In addition to Nep1 and Nep2, we discovered three additional nepenthesin homologues, which we name Nep3, Nep4, and Nep5 (Figure 4). Nep3−5 exhibit approximately 50% homology to each other as well as to Nep1 and Nep2, indicating they are distinct, novel aspartic proteases. Our preliminary sequencing of the N. ampullaria transcriptome revealed the presence of a Nep5 transcript that is 98% identical to N. rafflesiana Nep5 (Figure 4A and unpublished data), suggesting these novel nepenthesins may also be present in other Nepenthes species. Like Nep1 and Nep2, Nep3−5 all contain the conserved catalytic aspartate residues and the 12 cysteine residues involved in disulfide bridges (Figure 4B). The catalytic aspartate residues in these five nepenthesins, as well as those from other Nepenthes species, reside within two highly conserved regions: AIMDTGSDLWTGQ (aa110−123) and I3DSGTT (aa314−319; numbering based on Nep1 with the catalytic aspartate underlined). These regions are more variable and less extensive in other aspartic proteases, such as pepsin. Another distinguishing feature of the nepenthesins is the variable ∼20−24 aa nepenthesin-type aspartic protein (NAP)-specific region, which contains four of the conserved cysteine residues.9 Nep1−5 are different from the five aspartic proteases AP1−5 reported by An et al.,36 which contain a plant-specific insert of approximately 100 aa (instead of the short NAP-specific insert) commonly found in vacuolar plant aspartic proteases (vAP, Figure S2). Our N. rafflesiana transcriptome does contain several transcripts homologous to these vacuolar aspartic proteases (not shown), although we did not detect these vAPs in the Nepenthes digestive fluids. Nep3−5 appear to be present in the fluid in lower amounts than those of Nep1 and Nep2 (Table S5); however, the relative activity and contribution of Nep3−5 to the digestive process remain to be investigated.

We have also identified a novel class of prolyl-endoprotease in the fluid. We previously reported that the Nepenthes fluid can
Figure 5. Model for the mechanism of carnivory in *Nepenthes*. The enzymes identified in this study are highlighted in red, whereas enzymes suggested by other studies are in blue.15,23 Receptors lining the pitcher detect the presence of prey or pathogens, resulting in the secretion of digestive enzymes into the cavity of the pitcher. Glucanases and chitinases digest the cell wall and exoskeleton of captured prey and pathogens. Digestion of the prey proteins and nucleic acids by the proteases, nucleases, and phosphatases releases feedstocks for the nitrogen and phosphate requirements of the plant. Lipases, possibly from the microbiome, facilitate the digestion of lipids. The peroxidases may contribute to the host defense against pathogen by generating ROS. The ROS can also oxidize prey proteins, which may cause them to be more susceptible to degradation by proteases.

glycoside hydrolases are involved in the digestion of polysaccharides, such as chitin found in an insect exoskeleton, as well as in the cell wall of pathogens and plant debris. We discovered a protein containing a dopamine β-monooxygenase domain (DOMON, a heme- and sugar-binding motif found in some cytchrome and glycosyl hydrolases)12,31 as well as hevein, a chitin-binding domain associated with antifungal activity.44,45 Both are likely involved in aspects of glycan recognition and play a role in plant defense against pests and fungal infection. The two cationic peroxidases as well as a third peroxidase that we discovered may contribute to host defense against pathogens involving the generation of highly reactive oxygen species (ROS).

Finally, we discovered two Leu-rich repeat (LRR) proteins, one of which appears to be a receptor-like protein kinase. LRR is a common motif of 20–30 amino acids rich in leucines present in tandem repeats that mediate protein–protein interactions. LRR receptor-like protein kinases have been reported to be involved in plant development, germination, and pathogen resistance.36 Two of the remaining proteins identified correspond to the α and β subunits of ATP synthase/H+-ATPase and may function as a proton pump for acidification of the pitcher fluid.33
library using BLAST. The procedure used does not appear to retain a statistical foundation sufficient to assess the probability of the many proteins identified with only one or two peptides, and misidentifications are possible. For example, nepenthesins-2 is noted as identified with three tryptic fragments, but there are no K and R in the mature enzyme sequence in almost every species reported.

In our study, we have identified 36 proteins in a direct search of the cDNA library, which we prepared from RNA-seq data, ten of which were uniquely determined using the nonspecific digestion approach. It is clear that protein digestion represents the most aggressively targeted functionality in steady-state pitcher fluid. Five nepenthesins (three of which are novel: Nep 3–5) and two nepirosins (both novel: Npr1, Npr2) combine to provide a potent proteolytic digestion capacity to the fluid. Both represent intriguing departures from their core enzymatic classifications, as noted in recent studies from our laboratories. The newly discovered variants should prove valuable as reagents in proteomics applications. The relatively abundant acid-stable serine carboxypeptidase, in particular, suggests a useful laddering enzyme to aid in peptide sequencing and will be pursued.

We summarize the putative functions of the identified *Nepenthes* fluid proteins in a model for the mechanism of carnivory (Figure 5). The functions of the hydrolytic enzymes in the fluid are simple to infer. Glucanases and chitinases digest the cell wall and exoskeleton of captured prey and plant debris. Digestion of the prey proteins and nucleic acids by the proteases, nucleases, and phosphatases supply nitrogen and phosphate for absorption by the plant. Many of the hydrolytic enzymes involved in prey digestion also possess antibacterial and/or antifungal properties because they can digest pathogens and hence can be considered pathogenesis-related proteins themselves. The association between carnivory and pathogenesis is further supported by the presence of several specifically pathogenesis-related proteins such as the peroxidases, hevein (a chitin-binding protein), thamaunin-like protein, and LRR proteins. The peroxidases likely contribute to host defense against pathogens by generating ROS. It has been proposed that oxidation of prey proteins by peroxidase-generated ROS can enhance their susceptibility to degradation by proteases. Hevein domains bind chitin and display antifungal activity. Thamaunin-like protein likely serves a similar antifungal role, although we note that thamaunin is a highly potent natural sweetener, and this protein may serve as a chemoattractant to insect prey. The functions of the other proteins in the fluid are low abundance and more difficult to ascertain with confidence. The Leu-rich repeat (LRR) receptor-like protein kinase (RPK1) could stimulate the secretion of enzymes into the cavity of the pitcher upon sensing the presence of prey or pathogen, and the membrane-bound H+–ATPase may function as a proton pump for acidification of pitcher fluid. Both proteins are membrane-associated. They are somewhat surprising to be identified in the fluid but may represent shed domains. The remaining proteins are involved in energy metabolism, protein transport, protein expression, and synthesis (Table 1). They could represent intracellular plant proteins, because many of them lack a signal peptide, or elements of the microbial community that share stretches of sequence identity with entries in the *Nepenthes* transcriptome.

An overlapping set of proteins was observed by Schulze and colleagues in the acidic digestive fluid of the Venus flytrap, suggesting that these two different carnivorous plants use similar mechanisms for prey digestion. However, the major proteases in the Venus flytrap secretions are cysteine proteases, whereas the main proteases in *Nepenthes* are aspartic proteases. Cysteine protease activity has been suggested to be present in the *Nepenthes* fluid, as activity was inhibited by cysteine protease inhibitor E-64. Although numerous transcripts corresponding to putative cysteine proteases are present in our cDNA library, we did not detect any in the *Nepenthes* fluid. It remains possible that such determinations are species dependent; however, we note that a high degree of consistency has emerged in the identification of several enzymes across many varieties (aspartic proteases, glucanases, and chitinases, for example). The enzymatic properties do not appear to vary significantly.

**CONCLUSIONS**

Our improved proteomics workflow did not identify other proteins that have been reported to be present in the *Nepenthes* fluid and native to the plant. Pathogenesis-related protein 1 (PR-1) and an oxidoreductase were detected, but their scores were below our significance threshold. We could detect no lipases, xyllosidases, or galactosidases. The fluid represents a steady-state harvest; therefore, these enzymes are either transient members produced during fluid maturation or, more likely, were contributed by the plant microbiome. Although the acidity and antimicrobial properties of the *Nepenthes* fluid suppress the growth of pathogens, the fluid still hosts a diverse microbial community, which may form a mutualistic relationship with the host plant. A recent metagenomics study of the *Nepenthes* microbiome revealed as many as 18 bacteria phyla with proteobacteria representing the predominant class. This bacterial class was also found in our prefiltred transcriptome (Figure 2). Bacteria isolated from *Nepenthes* fluid have been reported to possess protease, chitinase, and lipase activities; however, it remains an open question whether these contribute meaningfully to mechanisms of carnivory. It is possible that bacterial colonization is mostly parasitic in nature; the topic requires further study. If the unusually broad and robust activity profile of the nepenthesins is any indication, the other acid-stable enzymes may present an equally broad substrate profile sufficient to supply amino acids, sugars, and phosphate from a range of sources.

**ASSOCIATED CONTENT**

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.6b00224.

Figure S1 provides an overview of the transcriptome analysis procedures in support of Figure 1; Figure S2 provides a phylogenetic analysis of nepenthesins relative to vacuolar aspartic proteases; and List S1 contains the set of amino acid sequences for the proteins identified in *N. rafflesiana* digestive fluid (PDF)

Table S1 provides a summary sequence analysis of the *N. rafflesiana* transcriptome; Table S2 provides a taxonomic analysis and assessment of the level of contaminations in the transcriptome; and Table S3 provides a summary of the *Nepenthes* transcriptome annotation all in support of
Figure 2; Tables S4–S6 provide details in support of the proteomics analyses summarized in Table 1 (XLSX)

**AUTHOR INFORMATION**

**Corresponding Author**

*Faculty of Medicine, University of Calgary, 3330 Hospital Drive NW, Calgary, Canada, T2N 4N1. Ph: 403-210-3811. E-mail: dschriem@ucalgary.ca.*

**Author Contributions**

This manuscript was written with contributions from all authors. All authors have approved the final version of the manuscript.

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

The work was supported by an NSERC Discovery Grant 298351-2010 to D.C.S., who acknowledges the additional support of the Canada Research Chair program, Alberta Ingenuity—Health Solutions and the Canada Foundation for Innovation.

**ABBREVIATIONS**

CDS, coding sequence; CID, collisionally induced dissociation; CTAB, cetyltrimethylammonium bromide; EDTA, ethylene diamine tetracetate; EMPI, exponentially modified protein abundance index; ETD, electron transfer dissociation; FASP, filter-aided sample preparation; LC–MS/MS, liquid chromatography tandem mass spectrometry; MS, mass spectrometry; Nep, nepenthesin; Npr, neprosin; ORF, open reading frame

**REFERENCES**


(7) Moran, J. A.; Clarke, C.; Greenwood, M.; Chin, L. Tuning of color contrast signals to visual sensitivity maxima of tree shrews by three Bornean highland Nepenthes species. *Plant Signaling Behav.* 2012, 7 (10), 1267–70.


(49) Stintzi, A.; Heitz, T.; Prasad, V.; Wiedemann-Merdinglu, S.; Kaufmann, S.; Geoffroy, P.; Legrand, M.; Fritig, B. Plant ‘patho-