1 **Understanding *K. phaffii* (*Pichia pastoris*) Host Cell Proteins: Proteomic Analysis and Flow-through**
2 **Affinity Clearance**

3 Sobhana A. Sripada[1], Driss Elhanafi[2], Leonard Collins[3], Taufika I. Williams[3], Marina Linova[4], John Woodley[4],
4 Stefano Menegatti[1,2]

5 1. Department of Chemical and Biomolecular Engineering, North Carolina State University, 911 Partners
6 Way, Raleigh, NC 27695, USA
7 2. Biomanufacturing Training and Education Center (BTEC), 850 Oval Drive, Raleigh, NC 27606, USA
8 3. Molecular Education, Technology, and Research Innovation Center (METRIC), North Carolina State
9 University, 2620 Yarbrough Dr., Raleigh, NC 27607, USA
10 4. Department of Chemical and Biochemical Engineering, Technical University of Denmark, 240, 2800
11 Kongens Lyngby, Denmark
12 Corresponding author: smenega@ncsu.edu.

13

14 **Abstract.** *K. phaffii* is a versatile expression system that is increasingly utilized to produce biological ther-
15 apeutics – including enzymes, engineered antibodies, and gene-editing tools – that feature multiple sub-
16 units and complex post-translational modifications. Two major roadblocks limit the adoption of *K. phaffii*
17 in industrial biomanufacturing: its proteome, while known, has not been linked to downstream process
18 operations and detailed knowledge is missing on problematic host cell proteins (HCPs) that endanger pa-
19 tient safety or product stability; furthermore, the purification toolbox has not evolved beyond the capture
20 of monospecific antibodies, and few solutions are available for engineered antibody fragments and other
21 protein therapeutics. To unlock the potential of yeast-based biopharmaceutical manufacturing, this study
22 presents *(i)* a secretome survey of *K. phaffii* cell culture harvests that highlights HCPs with predicted im-
23 munogenicity, ability to cause product instability by proteolysis or degradation of excipients, and potential
24 to interfere with purification operations via product association or co-elution; and *(ii)* a novel affinity ad-
25 sorbent functionalized with peptide ligands that target the whole spectrum of *K. phaffii* HCPs – Pich-
26 iaGuard – designed for the enrichment of therapeutic proteins in flow-through mode. The PichiaGuard
27 adsorbent features high HCP binding capacity (~25 g per liter of resin) and successfully purified a mono-
28 clonal antibody and an ScFv fragment from clarified *K. phaffii* harvests, affording up to 80% product yield,
29 and a >300-fold removal of HCPs. Notably, PichiaGuard outperformed commercial ion exchange and
30 mixed-mode resins in removing high-risk HCPs – including aspartic proteases, ribosomal subunits, and
31 other peptidases – thus demonstrating its value in modern biopharmaceutical processing.

32

35

36 **1. Introduction.** The engineering of cell lines in the *Pichia* genus has been pursued for more than two
37 decades – by improving the design of plasmid, promoters, and signal peptides – to achieve high produc-
38 tivity and fidelity of high-value therapeutics. In particular, the species *Komagataella phaffii (K. phaffii)* is
39 currently utilized as an expression host for a variety of proteins including antibodies – whole monoclonals
40 and engineered fragments – as well as fusion proteins, virus-like-particles, etc. [1–3]. These efforts have
41 been conducted in concert with system biology to better understand *K. phaffii*'s genetic repertoire as well
42 as the temporal evolution of transcriptomics and proteomics across the various phases of cell growth
43 under different conditions [4–12]. Nonetheless, a recent SWOT analysis on the adoption of *K. phaffii* as a
44 platform for protein production highlighted the need to intensity efforts towards secretion rate and re-
45 ducing protease-mediated product degradation. While capable of both intra- and extra-cellular protein
46 production, *K. phaffii* features an innately slow secretory trafficking [13,14]. These issues can be mitigated

47 by improving gene regulation [15][16] or developing strains with lower proteases content [17], although
48 these modifications can negatively impact cell growth rates and viability [18]. Therefore, cell engineering
49 and culture efforts must be adjuvated by purification technologies capable of removing *K. phaffii* host cell
50 proteins (HCPs) with enzymatic and/or immunogenic activity as early as possible in the manufacturing
51 pipeline [18].

52 Current biotherapeutic formulations contain a final average of ~20 ng HCP per mg drug product, typ-
53 ically represented by less than ten species [19],[20]. Although *K. phaffii* naturally produces fewer HCPs
54 than its mammalian counterparts, failure to reduce their titer below a prescribed limit puts at risk the
55 stability of the product and the safety of the patient. Decades of production of therapeutic monoclonal
56 antibodies (mAbs) in Chinese hamster ovary (CHO) cells has pointed at the presence of HCPs that conju-
57 gate an inherent toxic or immunogenic activity with the ability to escape purification by coeluting with
58 the mAb product. Recently, research groups have assessed the harmful potential of HCPs secreted by non-
59 mammalian and microbial cell lines employed – of forthcoming – in biopharmaceutical manufacturing
60 [21,22]. However, the identity and values of safety threshold of *K. phaffii* HCPs and their ability to evade
61 clearance by current chromatographic technology has been but barely investigated. For example, recent
62 *K. phaffii* engineering efforts have been associated with increased co-expression of isomerases and pro-
63 teases whose CHO homologues are known as immunogenic and difficult-to-remove, although their po-
64 tential interference with downstream processing is as of yet unknown [15,16,23].

65 Current processes for the purification of *K. phaffii*-secreted proteins rely on a train of operations in-
66 cluding cell separation, multiple chromatographic steps, and final concentration or filtration. As in the
67 CHO pipeline, the chromatographic segment typically comprises an affinity-based capture step followed
68 by polishing via ion exchange and hydrophobic interaction modalities (**Table S1**) [23]. As these cells are
69 harvested at high densities, a liter of culture fluid may contain up to a gram of HCPs, whose distribution
70 of physicochemical and biochemical properties varies with the target product and culture conditions. Such
71 process- and product-dependent diversity of the *K. phaffii* secretome, combined with the current dearth
72 of knowledge of its impact on purification technology and – most importantly – the lack of affinity tech-
73 nology dedicated to the removal of HCPs increases the risk of contaminants present in polished pools.

74 Developing a true platform process, capable of sustaining the diversity in titer and biomolecular make-
75 up of industrial feedstocks, can benefit significantly by introducing an affinity resin with robust HCP and
76 hcDNA capture activity. In prior work, we introduced the paradigm of "flow-through affinity chromatog-
77 raphy" and developed an adsorbent – named LigaGuard™ [24] – functionalized with peptide ligands de-
78 signed to target the whole CHO proteome spectrum. Our team demonstrated LigaGuard™ by purifying
79 therapeutic mAbs from an ensemble of seven CHO harvests: when utilized alone, LigaGuard™ afforded a
80 logarithmic HCPs removal value (HCP LRV) ranging between 1.5-2.5; when coupled with an affinity resin
81 operated in bind-and-elute mode, the two-step process afforded cumulative HCP and hcDNA LRVs ~ 4,
82 corresponding to a final HCP level of 8 ppm. Additionally, LigaGuard™ managed to clear high-risk and hard-
83 to-remove HCPs – such as Cathepsins and Phospholipases – thus demonstrating it potential to de-risk and
84 simplify the chromatographic pipeline of protein purification.

85 In this study, we sought to expand our study to yeast expression systems by conducting a bioprocess-
86 targeted proteomic survey of *K. phaffii* cell culture harvests, proposing a mechanism for identifying and
87 annotating HR-HCPs and by developing an adsorbent – named PichiaGuard – for the removal of *K. phaffii*
88 HCPs and hcDNA via flow-through affinity chromatography. In the first part of our study, we identified
89 and characterized common HCPs in the *K. phaffii* secretome and evaluated their risk and immunogenicity
90 profiles using bioinformatics tools. We then leveraged the resultant biomolecular information to design
91 and screen a peptide library against the *K. phaffii* secretome to identify selective HCP-targeting peptide
92 ligands. We utilized the discovered ligands to formulate PichiaGuard, which was evaluated in terms of HCP
93 binding capacity and purification for monoclonal full mAb and a ScFv fragment from clarified *K. phaffii*
94 fluids. A proteomics analysis of the effluents was conducted to evaluate the HCP capture performance of

95 the peptide-based resins. Globally, our results indicate that PichiaGuard outperforms state-of-the-art
96 chromatographic resins in terms of both capture of *K. phaffii* HCPs – notably, among them, aspartic pro-
97 teases, peptidases, ribosomal subunit proteins, heat shock proteins etc. – and product recovery, demon-
98 strating its value for the purification of high-value biotherapeutics produced by yeast hosts.
99
100

101 **2. Materials and Methods**

102

103 *2.1. Materials.* Fmoc-protected amino acids Fmoc-Gly-OH, Fmoc-Ser(tBu)-OH, Fmoc-Ile-OH, Fmoc-Ala-
104 OH, Fmoc-Phe-OH, Fmoc-Tyr(tBu)-OH, Fmoc-Asp(OtBu)-OH, Fmoc-His(Trt)-OH, Fmoc-Arg(Pbf)-OH, Fmoc-
105 Lys(Boc)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Glu(OtBu)-OH, Fmoc-Pro-OH, Fmoc-Trp(Boc)-OH, Fmoc-Cys(Trt)-
106 OH, and Fmoc-Leu-OH, the coupling agent azabenzotriazole tetramethyl uronium hexafluorophosphate
107 (HATU), diisopropylethylamine (DIPEA), piperidine, and trifluoroacetic acid (TFA) were sourced from
108 ChemImpex International (Wood Dale, IL, USA). The Toyopearl AF-Amino-650M (PichiaGuard TP-650M)
109 and HW-50F (PichiaGuard TP-50F) resins were obtained from Tosoh Bioscience (Tokyo, Japan). Phosphate
110 buffered saline at pH 7.4, triisopropylsilane (TIPS), 1,2-ethanedithiol (EDT), anisole, Protease Inhibitor
111 Cocktail, Kaiser test kits, and 3 kDa and 10 kDa MWCO centrifugal filters were from Millipore Sigma (St.
112 Louis, MO, USA). N,N'-dimethylformamide (DMF), dichloromethane (DCM), methanol, N-methyl-2-pyrrol-
113 idone (NMP), sodium phosphate (monobasic), sodium phosphate (dibasic), hydrochloric acid, sodium ac-
114 etate, glacial acetic acid, Formic acid, Tris-HCl at pH 7.5 and 8, Trypsin (100μg), DTT, 660 nm Histidine-rich
115 protein quantification and bicinchoninic acid (BCA) assay reagents were obtained from Fisher Chemicals
116 (Hampton, NH, USA). Human IgG was obtained from Athens Bio (Atlanta, GA). Ammonium Bicarbonate
117 was sourced from Acros, Sodium Deoxychlate, IAA and CaCl2 from Sigma Aldrich and Zwittergent 3-16
118 was obtained from CalBioChem. Vici Jour PEEK chromatography columns (2.1 mm ID, 30 mm length, 0.1
119 mL volume) and polyethylene frits were obtained from VWR international (Radnor, PA, USA). The 10-20%
120 Tris-Glycine HCl SDS-PAGE gels, Urea and Coomassie blue stain were purchased from Bio-Rad Life Sciences
121 (Carlsbad, CA, USA). Alexa Fluor 488 and 594 dyes, and Pierce dye removal columns were obtained from
122 Thermo Fisher Scientific. HiPrep Desalting 26/10 column and Sepharose CM ion exchange resin were ob-
123 tained from Cytiva.

124

125 *2.2. Proteomics analysis.* A label-free relative quantification workflow was developed to analyze the K.
126 phaffii secretome using a bottom-up approach. Filter-aided sample preparation (FASP) [25] was per-
127 formed with 30 kDa MWCO filtration units, each passivated with 20 μL of 0.1M TRIS-HCl pH 8 prior to
128 analyses. The starting volume for each sample was 200 μL (with varying protein amounts per sample) and
129 the BCA assay was used to determine total protein amount in each sample volume. Addition of 15 μL of
130 50 mM dithiothreitol or DTT (in 0.1 M TRIS-HCl at pH 8) to each sample - with incubation for 30 minutes
131 at 56°C - served to cleave disulfide bonds in sample proteins. This was followed by further protein dena-
132 turation by the addition of 200 μL of 8 M urea (in 0.1 M TRIS-HCl at pH 8) to each sample, with vortex and
133 spinning down. Each sample was next transferred onto a separate, passivated, 30 kDa MWCO filtration
134 unit for FASP. The samples were then concentrated by centrifugation at 12,000g for 15 minutes at 21°C
135 and the flow-through was discarded. An additional 200 μL of 8 M urea was added to each filtration unit,
136 immediately followed by the addition of 64 μL of 200 mM iodoacetamide or IAA (in 8 M urea). Each MWCO
137 filtration unit now carried 264 μL of solution (containing 50 mM IAA). Incubation for 60 minutes at room
138 temperature in the dark ensured the alkylation of free sulfhydryl groups of cysteine in the sample pro-
139 teins. This was followed by centrifugation at 12,000g for 15 minutes at 21°C. Subsequent washing steps
140 were performed, involving multiple rounds of centrifugation (12,000g for 15 minutes at 21°C) and resus-
141 pension into TRIS-based buffers (3 rounds with 2 M Urea with 10 mM CaCl$_2$ in 0.1 M TRIS-HCl at pH 8 and
142 3 rounds with the digestion buffer 0.1 M TRIS-HCl pH 7.5) before the addition of trypsin to the MWCO

143 filters (1:50 enzyme/protein ratio – based on BCA assay results – in 0.1 M TRIS-HCl pH 7.5) for overnight
144 proteolytic digestion. All flow-through solutions up to this point were discarded.

145 The overnight digestion was disrupted using a solution of 0.001% v/v Zwittergent with 1% v/v
146 formic acid solution or quench buffer. This quench buffer was added in steps (with centrifugation at
147 12,000g for 15 minutes at 21°C in between steps, combining and collecting the flow-through solutions
148 from a given sample) up to a total volume of 450 μL per sample. Each sample, now comprised of 450 μL
149 of tryptic peptide solution, was lyophilized and stored at -20°C until nanoLC-MS/MS. Upon reconstitution
150 in 200 μL of Mobile Phase A (MPA: 98% v/v water, 1.9% v/v acetonitrile, 0.1% v/v formic acid), reverse-
151 phase separation of the tryptic peptide fragments was carried out using a PepMap 100 C18 trap column
152 (3 μm particle size, 75 μm ID, 20 mm length) in series with an EASY-Spray C18 analytical column (2 μm
153 particle size, 75 μm ID, 250 mm length) - in a 'trap-and-elute' configuration - installed on an EASY nanoLC-
154 1200 instrument (Thermo Scientific, San Jose, CA) interfaced with an Orbitrap-Exploris-480 (Thermo Sci-
155 entific, Bremen, Germany). The flowrate was maintained at 300 nL/min. Peptides were eluted using a 105
156 minutes solvent gradient, ramping from 5% to 25% Mobile Phase B (MPB: 80% v/v acetonitrile, 19.9% v/v
157 water, 0.1% formic acid) over 75 min, followed by another ramp to 40% MPB over 10 min, and then a
158 steep ramp to 95% MPB in 1 min, at which point MPB was maintained at 95% for 17 minutes for column
159 washing. Eluted tryptic peptides were ionized by subjecting them to 1.9 kV in the ion source for elec-
160 trospray ionization. The ion transfer tube temperature was maintained at 275°C. The peptides were inter-
161 rogated by full MS scan and data-dependent acquisition (DDA) MS/MS. DDA was performed as a 105-
162 minute nanoLC-MS/MS method. Scan parameters for full MS were: 120,000 resolution, RF Lens of 40%,
163 maximum injection time of 120 ms, scan range of 375-1,600 m/z, with normalized AGC setting at 300%.
164 Dynamic exclusion (with 20 s exclusion duration) was used to minimize re-interrogation of pre-sampled
165 precursors. Cycle time was set at 1.5 s. Charge states 2-6 were included in the analysis. An intensity thresh-
166 old of 1.5×10^4 was set for $ddMS^2$. Scan parameters for $ddMS^2$ were: 15,000 resolution, inclusion window
167 of 1.5 m/z, maximum injection time of 21 ms, HCD collision energy of 30% (fixed), with normalized AGC
168 setting at 100%. Quality control in nanoLC-MS/MS analyses involved blank runs, as well as standard BSA
169 digest and standard HeLa digest runs at regular intervals within the length of the sample queue.

170 Post-acquisition data analysis and automated LFQ was performed by Proteome Discoverer v. 2.4
171 (PD, Thermo Scientific, San Jose, CA) with appropriate grouping variables. The raw nanoLC-MS/MS files
172 were interrogated against the *Pichia pastoris* / *Komagataella phaffii* protein database (4,983 sequences)
173 downloaded from UniProt. The data was also searched against a contaminants database (69 sequences)
174 to identify potential contaminants during the experiments. These databases were searched with the fol-
175 lowing parameters: trypsin (full) as the digesting enzyme, a maximum of 3 missed trypsin cleavage sites
176 allowed, 5 ppm precursor mass tolerance, 0.02 Da fragment mass tolerance, dynamic modifications on
177 methionine (oxidation), N-terminus (acetyl), methionine (met-loss+acetyl), methionine (met-loss), as well
178 as static carbamidomethyl modifications on cysteine residues. The SEQUEST HT algorithm was employed
179 in data interrogation. This algorithm compares the observed peptide MS/MS spectra and theoretically
180 derived spectra from the target database(s) to assign appropriate quality scores. These scores and other
181 important predictors are combined in the algorithm, which assigns a composite score to each peptide.
182 The overall confidence in protein identification is increased with increasing number of distinct amino acid
183 sequences identified. Therefore, proteins are normally categorized into a different priority group depend-
184 ing on whether they have only one or multiple unique tryptic peptide sequences of the required peptide
185 identification confidence. Percolator peptide validation was based on the q-value (adjusted p-value) and
186 minimal false discovery rate (FDR) < 0.01 was selected as a condition for successful tryptic peptide assign-
187 ments.

188

189 ***2.3. Design of K. phaffii strain X-33 producing ScFv13R4****. The DNA sequence encoding the ScFv13R4 pro-
190 tein [26] was first codon-optimized for efficient expression in *Pichia pastoris* wild type X-33 (Life

191 Technologies Ltd, Grand Island, NY). The gene was amplified via PCR using primers scFv13R4 αF1
192 (GCATGAATTCATGGCAGAAGTCCAATTAG) and scFv13R4 αR1 (GAATGCGGCCGCTCATAA-
193 GACTGCCAACTTAGTACC) carrying the extra restriction sites EcoRI and NotI (New England Biolabs, Beverly,
194 MA), respectively. The PCR product was cloned in the vector pPICZα-A (Life Technologies Ltd, Grand Island,
195 NY) at EcoRI and NotI sites generating a fusion protein ScFv13R4 to both His and C-myc tags present on
196 the expression vector. After electroporation in *E. coli* Stbl4 and selection on LB agar plates supplemented
197 with zeocin at 25 µg/mL, positive recombinant plasmids were identified using colony PCR. One recombi-
198 nant plasmid was linearized with the restriction enzyme SacI and introduced into *Pichia pastoris* X-33 via
199 electroporation. Selection was performed on YPD plates supplemented with zeocin at a concentration of
200 1000 µg/mL and positive clones were screened with colony PCR.
201
202 **2.4. Production of K. phaffii WT (null-X-33) and ScFv13R4 cultures**. Overnight culture was initiated from
203 a single colony in 3 mL of YPD medium and incubated at 30°C under agitation at 350 rpm. A volume of 1
204 mL of this culture was used as the inoculum to initiate a new 50 mL culture in YPD medium, which was
205 grown at the same conditions for 24 hours. In order to adapt the cells to the production medium, cells
206 were centrifuged at 3000g for 10 minutes and resuspended in 50 mL of salt media BFM21 containing
207 glycerol at 40 g/L (3.5 mL of 85% H_3PO_4, 0.119 g $CaSO_4$, 2.4 g K_2SO_4, 1.950 g $MgSO_4$:$7H_2O$, 0.650 g KOH,
208 0.940 g sodium citrate, 5.0 g ammonium sulfate, 1.537 g NaOH, 31.75 mL glycerol, 40 mL 1.0 M acetate
209 buffer, 0.25 mL antifoam 240, pH 5.0) and supplemented with trace elements PTM4 (12 ml/L: 1275 µL of
210 H_2SO_4, 0.5 g $CuSO_4$:$5H_2O$, 0.020 g NaI, 0.750 g $MnSO_4$, 0.05 g Na_2MoO_4:$2H_2O$, 0.005 g H_3BO_3, 0.125 g $CoCl_2$,
211 1.675 g $ZnCl_2$, 5.4 g $FeSO_4$, 0.050g added per 250 mL of medium). Cell growth was allowed to proceed for
212 24 hours at 30°C under agitation at 350 rpm. This culture was in turn used to inoculate a 1-L working
213 volume bioreactor containing BFM21 supplemented with PTM4 (12 ml/L) medium and glycerol (40 g/L)
214 starting at low optical density (OD_{550nm} ~ 2). Fermentation was conducted in multiple bioreactors using
215 the following parameters: pH controlled at 5 using 28% v/v ammonium hydroxide and 80% v/v phosphoric
216 acid; temperature at 30°C; and dissolved oxygen maintained at 30% in cascade with agitation. The initial
217 glycerol was exhausted within 24 hours of fermentation, after which a glycerol feed phase was initiated,
218 and the fermentation was continued until the fluid reached an OD_{550nm} ~ 100. At this point, the suspension
219 from one reactor was collected and centrifuged at 4000g for 30 minutes to remove the wet cell mass,
220 while ensuring minimal cellular rupture and intracellular protein release. Simultaneously, a methanol feed
221 phase was started in the other bioreactor by adding it at a concentration of 0.1% v/v at 0.3 mL/min while
222 maintaining the dissolved oxygen at 30%, to induce ScFv13R4 production. After 48 hours of fermentation
223 in methanol and upon achieving a cell density of OD_{550nm} ~ 300, the cells were harvested via centrifugation
224 at 3000g for 10 minutes. The supernatant was then filtered through a 0.45 µm filter before further use.
225
226 **2.5. Peptide synthesis.** A tetrameric one-bed one-peptide (OBOP) library in the format X_1-X_2-X_3-X_4-GSG
227 was prepared on HMBA-ChemMatrix resin via direct solid-phase Fmoc/tBu peptide synthesis using a Syro
228 I automated peptide synthesizer (Biotage, Uppsala, Sweden), as described in prior work [24,27]. Amina-
229 tion of Toyopearl HW-50F resin was performed by incubating 1 g of resin with 2 mmol of carbonyldiimid-
230 azole in 30 mL of acetone for 30 minutes at 25°C under mild agitation, followed by rinsing in acetone and
231 incubation with 2 mmol of tris(2-aminoethyl) amine under the same conditions, to obtain an amine den-
232 sity of ~0.2 mmol per gram of resin. The peptide sequences were synthesized on Toyopearl AF-Amino-
233 650M (TP650M) and aminated HW-50F resins (TP50F) following the same protocol using an Alstra auto-
234 mated peptide synthesizer (Biotage, Uppsala, Sweden). Upon completing chain elongation, the peptides
235 were deprotected using TIPS deprotection cocktail (95% v/v TFA, 2.5% v/v TIPS, 2.5% v/v water). The Pich-
236 iaGuard-TP650M and PichiaGuard-TP50F adsorbents were formulated by combining equal volumes of the
237 above-listed peptide-functionalized resins.
238

239 **2.6. Protein labeling and formulation of a screening mix.** Human IgG and the HCPs contained in the *P.*
240 *pastoris* (X-33 strain) cell culture fluid and were labeled using NHS ester AlexaFluor dyes AF488 (green)
241 and AF594 (red), respectively. Specifically, each dye was initially dissolved in dry DMSO at the concentra-
242 tion of 10 mg/mL. Following diafiltration of the X-33 cell culture fluid into PBS at pH 7.4 using 3 kDa MWCO
243 filters, 100 μL of *K. phaffii* HCPs at the concentration of 0.5 mg/mL were incubated with 1 μL of AF594; in
244 parallel, 100 μL of a solution of IgG at 5 mg/mL in PBS at pH 7.4 was incubated with 1 μL of AF594. The
245 labeling reactions were allowed to proceed for 3 hours at 25°C under mild agitation. Spin columns were
246 used to remove the unreacted dyes. The labeled proteins were then mixed to obtain a screening mix
247 comprising AF488-IgG at 2 mg/mL and AF594-HCPs at 0.2 mg/mL. A volume of 20 μL of OBOP library beads
248 were then incubated with 1 mL of screening mix overnight at 4°C under mild agitation and washed with
249 PBS before proceeding with the screening protocol.

250

251 **2.7. Library screening.** The OBOP tetrameric library was screened against the screening mix (see *Section*
252 *2.6)* using a microfluidic device developed by our team [28]. Briefly, the microfluidic device comprises an
253 imaging chamber which hosts individual library beads in rapid sequence and is installed in a DMi8 micro-
254 scope (Leica, Wetzlar, Germany) equipped with an ORCA-Flash4.0 V3 Digital CMOS camera with W View
255 Gemini image splitting optics (Hamamatsu, Shizuoka, Japan). A MATLAB® GUI developed to operate this
256 screening system was implemented to visualize the library beads under the green (human IgG; excita-
257 tion/emission: 488/496 nm) and red (*K. phaffii* HCPs; excitation/emission: 594/617 nm) channels. Library
258 beads with green-only or green-and-red fluorescent emission were discarded, whereas beads with strong
259 red-only fluorescent emission were collected as positive leads and analyzed via Edman Degradation using
260 a PPSQ 33-A Protein Sequencer (Shimadzu, Kyoto, Japan) to identify the peptide sequences carried
261 thereon.

262

263 **2.8. Diafiltration of feedstocks.** Prior to chromatographic purification, the clarified X-33 cell culture har-
264 vests were diafiltered to remove small media component, chiefly glycerol, and unreacted protease inhib-
265 itors as well as adjust the conductivity and pH using a HiPrep™ 26/10 column (Cytiva, Marlborough, MA)
266 into two different buffers, namely 20 mM sodium acetate at pH 5.7- and 25-mM sodium phosphate at pH
267 7.2. The HCP titers were varied respectively between 0.3 mg/mL and 0.5 mg/mL in both the acetate-con-
268 ditioned and phosphate-conditioned feedstocks as estimated by BCA.

269

270 **2.9. Static and dynamic HCP binding capacity measurements.** Static binding experiments were conducted
271 by incubating aliquots of 30 μL of settled PichiaGuard-TP650M or PichiaGuard-TP50F resin in a 12-well
272 plate (WP) with acetate-conditioned and phosphate-conditioned X-33 feedstocks featuring HCP titers of
273 $0 - 1$ mg/mL for 3.5 hours at 25°C under mild agitation (300 rpm). The plate was then spun down at 3000
274 rpm for 10 minutes and the supernatants were collected and analyzed to determine the residual HCP titer
275 via BCA. Finally, the values of static binding capacity were calculated via mass balance. The dynamic bind-
276 ing capacity was measured via breakthrough experiments ($DBC_{20\%}$) by loading ~6 mL of acetate-condi-
277 tioned or phosphate-conditioned X-33 feedstocks featuring a HCP titer of ~0.4 mg/mL onto 0.1 mL column
278 packed with chromatographic resin – namely, PichiaGuard-TP650M and PichiaGuard-TP50F resin, Capto
279 Q or CaptoAdhere (Cytiva), or CHO LigaGuard™ (LigaTrap Technologies, Raleigh, NC) – pre-equilibrated
280 with the respective buffers – namely, 20 mM sodium acetate at pH 5.7 and 25 mM sodium phosphate at
281 pH 7.2 – at a residence time (RT) of 1 minute until saturation was reached. The effluent was continuously
282 monitored via spectrophotometry at 280 nm and apportioned in 0.5 mL fractions that were finally ana-
283 lyzed to determine the residual HCP titer via BCA.

284

285 **2.10. Purification of ScFv and mAb via flow-through affinity chromatography using PichiaGuard.** Pich-
286 iaGuard resin was initially packed in a 0.1 mL column and equilibrated with 20 mM sodium acetate at pH

287 5.7 at 0.2mL/min for 15 minutes. Two *K. phaffi* harvests were utilized as feedstock in this study – the first
288 one featuring a ScFv titer of ~1 mg/L and an HCP titer of ~0.6 mg/mL, while the second one featuring a
289 mAb titer of ~0.4 mg/mL and an HCP titer of ~0.5 mg/mL – both conditioned in 20 mM sodium acetate at
290 pH 5.7. A volume of 4 mL of harvest was then loaded on the column at the residence time (RT) of 1 minute
291 and flow-through fractions of 0.3 mL were collected throughout the load and final column wash for ana-
292 lytical characterization (*Sections 2.2, 2.11, and 2.12*). All purification studies were performed using an
293 ÄKTA pure (Cytiva, Chicago, IL, USA) while monitoring the effluents using UV spectroscopy at 280 nm.
294
295 ***2.11. Reverse phase (C18 RP-HPLC) analysis for the quantification of ScFv13R4 titer and purity.*** The titer
296 of ScFv13R4 in the collected chromatographic fractions was measured via C18 RP-HPLC using a Zorbax
297 C18-SB column (Agilent, Santa Clara, CA) installed on a Waters Alliance 2690 HPLC. The column was equil-
298 ibrated by flowing 95% mobile phase B (MPB: 0.5% v/v TFA in acetonitrile) and 5% mobile phase A (MPA:
299 0.5% v/v TFA in water), at a flowrate of 0.5 mL/min. A 20-minutes gradient method (95 – 45% MPB) was
300 conducted at the constant flow rate of 0.5 mL/min to achieve separation. Pure ScFv13R4 expressed in *E.*
301 *coli* cells was utilized to identify the product retention time and calculate the ScFv13R4 purity in the chro-
302 matographic fractions.
303
304 ***2.12. Other Analytical Assays****.* Total protein and *P. pastoris* HCP titers in control studies were measured
305 via BCA assay (Pierce BCA by ThermoFisher, NC) and *Pichia pastoris* HCP ELISA kit (Generation 2, Cygnus
306 Technologies, SC) following the manufacturer's protocol. Product purity was evaluated using 660 nm BCA
307 for histidine rich proteins and SDS-PAGE analysis under non-reducing conditions. mAb was quantified us-
308 ing an analytical proteinA method as described previously [24].
309
310
311 **3. Results**
312
313 ***3.1. Proteomics analysis and immunogenicity annotation of the K. phaffii secretome****.* The remarkable
314 progresses in mass spectrometry and statistical correlation models achieved in the last decade have trans-
315 formed the proteomics of biological fluids from a complex endeavor to a routine procedure [29,30]. The
316 proteomic analysis of bioreactor harvests and bioprocess fractions has enabled the compilation of data-
317 bases that highlight features of host cell proteins (HCPs) that are relevant in biopharmaceutical manufac-
318 turing, such as the ability to associate with a therapeutic product [31], escape clearance by specific chro-
319 matographic modalities [32], enzymatic activity that causes product degradation or destabilization, and
320 immunogenicity or toxicity [33]. To date, the attention of both academic and industrial communities has
321 focused on the proteomics of Chinese hamster ovary (*i.e.*, *Crisetulus griseus*) cells, given their widespread
322 used in the production of monoclonal antibodies (mAbs), as exemplified by the databases CHOPPI and
323 CHOGenome [34,35].
324 The recent growth in use of yeast expression systems in biomanufacturing has highlighted the need
325 of producing analogous databases for *Pichia* hosts. Upon completing the sequencing of the full *K. phaffii*
326 genome, -omics approaches have indeed been implemented to acquire deeper understanding of the *K.*
327 *phaffii* expression system [36]. Proteomic studies conducted by various teams [4,37] have contributed to
328 establishing public data repositories – chiefly, the *Pichia* genome [38] – and have highlighted variations in
329 the expression and secretion of HCPs upon changes in feeding strategies (glucose *vs.* methanol), oxygen
330 levels during cell growth, feed composition, and location of expression (intra- *vs.* extra-cellular)
331 [5,11,39,40]. To date, up to 575 unique proteins have been identified in *K. phaffii*, with its secretome
332 consisting of at least 335 confirmed species [4,41]. These databases, however, do not contain annotation
333 on the risk profile of *K. phaffii* HCPs in the context of bioprocessing.

334   To overcome this gap, we sought to build a database of *K. phaffii* HCPs with *(i)* reported or predicted
335   immunogenicity, *(ii)* homology with CHO HCPs that have been shown to co-elute with mAbs from different
336   chromatographic modalities, and *(iii)* enzymatic activity that can compromise the function, structure, and
337   stability of recombinant products. To this end, we implemented a nanoLC-MS/MS workflow established
338   in prior work (**Figure S1A**, [42]) to conduct a bottom-up proteomic analysis of a supernatant produced by
339   culturing methylotropic wild-type (X-33) *K. phaffii* and identify HCPs that are secreted along with those
340   released by lysed cells. We note that bottom-up proteomics involves proteolytic (tryptic) digestion of bi-
341   oprocess samples and protein identification via reconstruction of primary sequences at the data-depend-
342   ent analysis (DDA) stage in nanoLC-MS/MS. While this may lead to the overestimation of certain functional
343   HCPs from pre-existing peptide fragments of HCPs cleaved in the cell culture, it also serves as a challenging
344   scenario for evaluation of novel purification technologies. The resultant spectral data was searched
345   against UniProt *K. phaffii* database (Proteome ID: UP000000314), which encompasses 5,073 proteins iden-
346   tified across three different *Pichia* species. The list of identified proteins and their UniProt IDs/RefSeq
347   accession numbers are listed in **Table S2**. Data interrogation indicated the presence of multiple proteases,
348   with an abundance of aspartic proteases that are especially active in acidic environments (*i.e.*, pH 3-5,
349   [43]), which is also a range frequently utilized for product elution at the affinity-based capture step [44].
350   Indeed, lower product yields caused by proteolytic degradation have often been observed with *K. phaffii*
351   [45], whose protease-rich secretome also results in a myriad of HCP fragments that can maintain their
352   immunogenic potential while also being challenging to remove due to non-covalent association to the
353   product. These proteins were identified and tagged 'risky' for our analyses.
354
355   In addition, an extended list of problematic *K. phaffii* HCPs was compiled by combining reports found
356   in the literature [4,18,23,41] with the estimated homology between *K. phaffii* HCPs and CHO HCPs with
357   known immunogenicity and a sequence-based 'Immunogenicity Potential' score predicted by the NIAID's
358   Immune Epitope Database. A list of 335 proteins identified in the clarified harvest (secretome) by Prote-
359   ome Discoverer, consistent with previously reported literature [41], were processed using the workflow
360   outlined in **Figure 1A**: the amino acid sequences of the *K. phaffii* HCPs identified via nanoLC-MS/MS were
361   aligned with a non-redundant database of *Cricetulus griseus* proteins (NCBI Tax ID: 10029) using the NCBI
362   BLASTp online tool by implementing a stringent definition of "homologous" hit, namely > 30% sequence
363   identity, an e-value < 10^{-10}, and a bit score > 50 based on the Pearson criteria. [46]. BLASTp returned 2,166
364   hits including multiple 'similar' sequence hits per each *K. phaffii* HCP. To avoid increasing redundancy, the
365   maximum number of hits per sequence was capped at 10 followed by application of the discussed criteria.
366   This list was then consolidated by 'summarizing' the sequence homology properties using the number of
367   BLAST hits per *K. phaffii* HCP as a metric, along with the properties mentioned above. In a data sampling
368   of homologous hits with > 50% sequence match, *K. phaffii* counterparts of peptidyl-prolyl isomerases,
369   protein disulfide isomerases and histones H2A/H3/H4, 40S/60S ribosomal proteins etc. have been identi-
370   fied – which have been associated with product aggregation/drug stability in CHO systems [32,33].
371
372   In parallel, the immunogenicity of *K. phaffii* HCPs was predicted using the MHC-II binding predictor
373   tool of IEDB by implementing a 'Consensus Approach' that estimates the propensity of a protein fragment
374   to promiscuously bind MHC-II alleles [47]. The sequences of *K. phaffii* HCPs were flanked as 15-mers and
375   tested for binding against 7-alleles (3 loci of HLA-DRB1, 2 loci of DRB3, 1 locus each of DRB4 and DRB5),
376   which have been observed to bind to antigenic peptides, leading to an immune response in 50% general
377   population (*note:* maximum allelic diversity has been observed with HLA-DRB1, whose 3 variants contrib-
378   ute to capture the diversity within this test set). The consensus approach for scoring MHC-II binding com-
379   bined the predictions from several algorithms (NN-align, SMM-align, Sturniolo/CombLib, NetMHCIIpan)
380   to ensure broad coverage and low rate of false negativity. The scored 15-mer fragments (n = 580,384,
381   (**Figure 1B**)) generated in this process were mapped back to the original *K. phaffii* HCP primary sequences

382   among which a subset of 'Agretope'-forming HCPs (*i.e.*, consensus percentile rank ≤ 1%, per IEDB recom-
383   mendations) were identified and flagged as 'risky' based on the ability of their fragments to bind more
384   than 2 tested allelic variants. As an example, thioredoxin peroxidase and peroxiredoxin-1 both belong to
385   the Peroxiredoxin family of enzymes – the latter of which has been reported as immunogenic in the CHO
386   proteome, and the former has been flagged by our workflow as immunogenic in *K. phaffii* [48].

387   We therefore utilized a combination of 'risky' proteases identified in the culture harvest, *K. phaffii*
388   HCPs with reported or predicted immunogenicity and homology to immunogenic CHO HCPs to identify a
389   set of 51 high-risk HCPs that have a potential for either immunogenicity or product degradation (**Figure
390   1C** and **Table 1**). Information regarding the HR-HCPs and their selection criteria (IEDB scores, Homology
391   parameters) are collated in **Table S2**. This ensemble of HR-HCPs is non-exhaustive and is meant to serve
392   as a guidance list for scientists developing therapeutic bioprocesses using *K. phaffii* hosts. We utilize this
393   database to track the persistence of bioprocess-relevant HCPs throughout the purification pipeline and
394   compare the performance of commercial *vs.* bespoke chromatographic adsorbents in terms of both global
395   HCP reduction and clearance of high-risk species through rest of the study.

396
397

398   **3.2. Identification of K. phaffii HCP-targeting peptide ligands.** The sequence-based bioinformatic analysis
399   of the *K. phaffii* HCPs identified in the harvest fluid presents a rather unique landscape of physicochemical
400   properties (**Figure S1B**; *note:* all values are calculated based on the amino acid sequence of the HCPs): *(i)*
401   the molecular weight distribution, centered at around 46.14 kDa, is significantly sharper (standard devia-
402   tion, σ ~29.03 kDa) and a narrower (7 – 193 kDa) compared to its CHO counterpart (σ ~ 35.91, 5 – 776
403   kDa, [24]); *(ii)* the values of isoelectric point feature a distinct bimodal distribution peaking at the values
404   of pH 5 and 9.5, with 40% of the species being anionic and 60% cationic at the physiological culture pH,
405   which represents a second sharp difference compared to CHO HCPs (61% anionic and 39% cationic); *(iii)*
406   the values of grand average hydropathy (GRAVY), varying between -1.84 and 0.49 (σ ~ 0.37), indicate that
407   *K. phaffii* HCPs are more hydrophilic than CHO HCPs (-1.90 to 0.97; σ ~ 0.26); and *(iv)* an average polarity
408   of 50.1% across all *K. phaffii* HCPs, with most species between a range of 42 – 62%, indicate a strong
409   propensity to form hydrogen-bond networks.

410

411   This analysis informed the design of a library whence peptides could be selected that target the whole
412   spectrum of the *K. phaffii* secretome: firstly, both positively and negatively charged amino acids (Glu, Lys,
413   His, Arg; Asp was excluded from the library due to the abundance of aspartyl proteases in *K. phaffii* har-
414   vests) were included to address anionic and cationic ensembles of HCPs; furthermore, given the abun-
415   dance of polar species, neutral amino acids that are hydrogen bond-forming (Gln and the spacer Ser) were
416   included to ensure the formation of hydrogen-bonding interactions; for the same reason, only one ali-
417   phatic amino acid (Val; Ile and Leu were excluded) and two aromatic amino acids capable of forming hy-
418   drogen-bonds (Tyr and Trp; Phe was excluded) were included; finally, two amino acid linkers, the flexible
419   Gly and the semi-rigid Ala (the rigid Pro was excluded) to sample the broadest possible range of confor-
420   mations. The multi-polar composition of the library was inspired by that of the ligands forming the CHO
421   LigaGuard™, among which multi-polar peptides hold a prominent role as key contributors to the capture
422   of high-risk and persistent CHO HCPs [27]. A short peptide length of 7 total residues in the format X_1-X_2-X_3-
423   X_4-G-S-G, wherein the 4 residues on the N-terminus form the HCP-binding segment and are varied within
424   the library, while the Gly-Ser-Gly tripeptide on the C-terminus acts as a flexible spacer to enhance ligand
425   display. This format was inspired by prior studies [49], where we observed that most of the energy of
426   protein:peptide binding is contributed by the 4 residues on the N-terminus, and the recommendation of
427   using PichiaGuard as a disposable – and ideally single-use – adsorbent, which requires moderate cost and
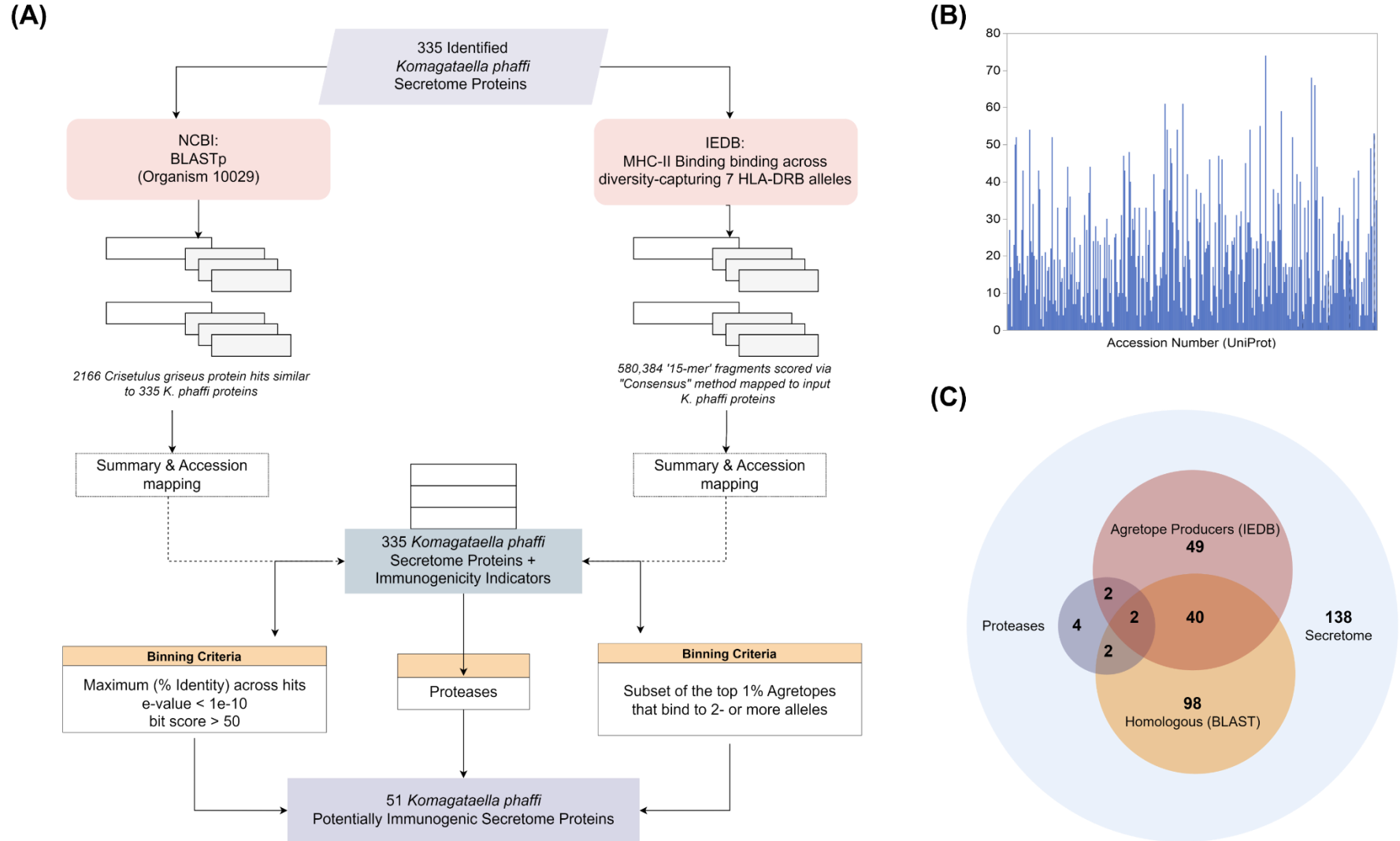428   thus short, easy-to-synthesize peptides.

Figure 1: **(A)** *Flowchart describing the identification and classification of high-risk K. phaffii host cell proteins (HR-HCPs) based on homology to CHO, immunogenicity designation (predicted by IEDB), and protease activity;* **(B)** *Histogram of 15-mer fragments generated per K. phaffii HCP identified in the cell culture feed: the fragments were utilized to score the cognate HCP using the IEDB MHC-II immunogenicity prediction tool;* **(C)** *Venn diagram showing the overlaps in the classification of HCPs among the various risk groups, which were ultimately chosen to compile the list of HR-HCPs presented in* **Table 1**.

434 *Table 1:* List of K. phaffii host cell proteins flagged as "high-risk" (HR-HCPs) and their homology to known CHO HR-HCPs. The high-risk classification includes: (1) HCPs frequently
435 encountered as problematic in bioprocessing, (2) HCPs with evidence for immunogenic potential, and (3) enzymatic proteases that can cause product degradation.

UniProt Accession #	Name	Homologous CHO HR-HCP	Gene Ontology: Molecular Function	Gene Name	*K. phaffii* HR Classification	Effects on Downstream Operations
C4QV16	Protein component of the large (60S) ribosomal subunit, nearly identical to Rpl4Ap		structural constituent of ribosome [GO:0003735]	PAS_chr1-3_0034	1,2	Recurrent observation in downstream processes
C4QV80	Dihydrolipoyllysine-residue succinyl transferase		dihydrolipoyllysine-residue succinyl transferase activity [GO:0004149]	PAS_chr1-3_0094	1,2	
C4QV89	Heat shock protein that cooperates with Ydj1p (Hsp40) and Ssa1p (Hsp70)		ATP binding [GO:0005524]; ATP hydrolysis activity [GO:0016887]	PAS_chr1-3_0102	1,2,3	Sorting and degradation of proteins
C4QV95	Ubiquitin-specific protease that deubiquitinates ubiquitin-protein moieties		thiol-dependent deubiquitinase [GO:0004843]	PAS_chr1-3_0108	1	Recurrent observation in downstream processes
C4QVY8	Translation initiation factor eIF4G, subunit of the mRNA cap-binding protein complex (EIF4F)		translation initiation factor activity [GO:0003743]	PAS_chr1-1_0053	1,2	
C4QY07	Phosphoglycerate kinase	Phosphoglycerate Kinase 1	ATP binding [GO:0005524]; phosphoglycerate kinase activity [GO:0004618]	PAS_chr1-4_0292	1,2	Recurrent observation in downstream processes
C4QY27	Subunit of the SWI/SNF chromatin remodeling complex			PAS_chr1-4_0309	1,2	
C4QZL4	Protein component of the large (60S) ribosomal subunit	60S Ribosomal Subunit Protein	RNA binding [GO:0003723]; structural constituent of ribosome [GO:0003735]	PAS_chr2-1_0087	1,2	Recurrent observation in downstream processes
C4QZQ5	Elongation factor 2		GTP binding [GO:0005525]; GTPase activity [GO:0003924]	PAS_chr2-1_0812	1,2	Recurrent observation in downstream processes
C4QZS3	Endoplasmic reticulum chaperone BiP	ER Chaperone BiP Precursor	ATP binding [GO:0005524]; ATP hydrolysis activity [GO:0016887]	PAS_chr2-1_0140	1,2	Impacts drug quality by increasing aggregation propensity
C4QZZ6	Dolichyl-phosphate-mannose--protein mannosyltransferase		dolichyl-phosphate-mannose-protein mannosyltransferase activity [GO:0004169]	PAS_chr2-1_0212	1,2	
C4R080	Proteasome endopeptidase complex	Alpha Enolase	proteasomal ubiquitin-independent protein catabolic process [GO:0010499]; ubiquitin-dependent protein catabolic process [GO:0006511]	PAS_chr2-1_0291	2	Leads to drug product modification via catalysis of dehydration reaction that converts glycerate to pyruvate
C4R080	Alpha 6 subunit of the 20S proteasome	proteasome subunit alpha type 7 isoform x1	proteolysis involved in protein catabolic process [GO:0051603]	PAS_chr2-1_0291	2	Recurrent observation in downstream processes
C4R095	Kex2 proprotein convertase		serine-type endopeptidase activity [GO:0004252]; serine-type endopeptidase inhibitor activity [GO:0004867]	PAS_chr2-1_0304	3	Signal peptidase for secretion, often overexpressed with recombinant gene
C4R0N8	Uncharacterized protein GN=PAS_chr2-1_0892 PE=4 SV=1		polyubiquitin modification-dependent protein binding [GO:0031593]	PAS_chr2-1_0892	1,2	
C4R0P1	Glyceraldehyde-3-phosphate dehydrogenase, isozyme 3	Glyceraldehyde-3 Phosphate Dehydrogenase	Enables oxidoreductase activity [GO:0016620]; NAD/NADP binding [GO:0050661]	PAS_chr2-1_0437	1	Recurrent observation in downstream processes

C4R0Q2	GTP-binding nuclear protein		GTP binding [GO:0005525]; GTPase activity [GO:0003924]	PAS_chr2-1_0449	1,2	Recurrent observation in downstream processes
C4R0Y3	Translationally controlled tumor protein homolog		cytoskeleton protein involved in translational activity [GO:0006412]	PAS_chr2-1_0522	1,2	
C4R142	Isocitrate dehydrogenase [NADP]		isocitrate dehydrogenase (NADP+) activity [GO:0004450]; magnesium ion binding [GO:0000287]; NAD binding [GO:0051287]	PAS_chr2-1_0580	2	
C4R146	Protein component of the small (40S) ribosomal subunit		structural constituent of ribosome [GO:0003735]	PAS_chr2-1_0584	2	
C4R1P7	Acetyl-coenzyme A synthetase	Glutaredoxin-2	acetate-CoA ligase activity [GO:0003987]; AMP binding [GO:0016208]; ATP binding [GO:0005524]	PAS_chr2-1_0767	1,2	Recurrent observation in downstream processes
C4R1P9	Pyruvate Kinase	Pyruvate Kinase PKM isoform X1	Nucleotide binding [GO:0000166] and transferase activity [GO:0016740]	PAS_chr2-1_0769	1,2	May lead to immunogenicity due to glycolytic action
C4R2G3	FK506-binding protein	Peptidyl-prolyl cis-trans isomerase	peptidyl-prolyl cis-trans isomerase activity [GO:0003755]	PAS_chr2-2_0476	1,2	Increases aggregation propensity by influencing folding and protein assembly in ER
C4R2H3	Thioredoxin peroxidase		thioredoxin peroxidase activity [GO:0008379]; unfolded protein binding [GO:0051082]	PAS_chr2-2_0220	1,2	
C4R2J7	Histone H3		DNA binding [GO:0003677]; protein heterodimerization activity [GO:0046982]	PAS_c034_0036 PAS_chr2-2_0199	1,2	Recurrent observation in downstream processes
C4R2S1	Catalase		catalase activity [GO:0004096]; heme binding [GO:0020037]; metal ion binding [GO:0046872]	PAS_chr2-2_0131	1,2	Recurrent observation in downstream processes
C4R306	Phosphomannomutase		phosphomannomutase activity [GO:0004615]	PAS_chr2-2_0053	2	
C4R3H8	Enolase I, a phosphopyruvate hydratase that catalyzes the conversion of 2-phosphoglycerate to phosphopyruvate	Beta-enolase isoform x3	Enables lyase activity [GO:0016829], enables phosphopyruvate hydratase activity [GO:0004634]	PAS_chr3_0082	2	Can lead to drug product modification via catalysis of dehydration reaction that converts glycerate to pyruvate
C4R3Q7	Peptidase A1 domain-containing protein		aspartic-type endopeptidase activity [GO:0004190]	PAS_chr3_1157	2,3	Aspartic protease
C4R3X8	ATPase involved in protein folding and the response to stress	Heat shock cognate protein	ATP binding [GO:0005524]; ATP hydrolysis activity [GO:0016887]	PAS_chr3_0230	1,2	Recurrent observation in downstream processes
C4R451	Aa-trans domain-containing protein	Vacuolar transporter	membrane protein [GO:0005774]	PAS_chr3_0295	1,2	
C4R458	Aspartic protease, attached to the plasma membrane via a glyco-sylphosphatidylinositol (GPI) anchor		aspartic-type endopeptidase activity [GO:0004190]	PAS_chr3_0303	2,3	Aspartic protease
C4R493	Subunit of the core complex of translation initiation factor 3(EIF3)		translation initiation factor activity [GO:0003743]	PAS_chr3_0340	1,2	
C4R4F5	Putative GPI-anchored aspartic protease		aspartic-type endopeptidase activity [GO:0004190]	PAS_chr3_0394	1,3	Aspartic protease

12

C4R4V6	Component of the RSC chromatin remodeling complex		zinc ion binding [GO:0008270]	PAS_chr3_0544	1,2	
C4R566	Phosphatidylinositol/phosphatidylcholine transfer protein		phosphatidylcholine transporter activity [GO:0008525]; phosphatidylinositol transfer activity [GO:0008526]	PAS_chr3_0655	1,2	
C4R5P4	Tetrameric phosphoglycerate mutase	Phosphoglycerate mutase 1	enables catalytic activity [GO:0004619]	PAS_chr3_0826	2	Recurrent observation in downstream processes
C4R5U7	S-adenosylmethionine synthase		ATP binding [GO:0005524]; metal ion binding [GO:0046872]; methionine adenosyltransferase activity [GO:0004478]	PAS_chr3_0876	1,2	
C4R626	Triosephosphate isomerase		triose-phosphate isomerase activity [GO:0004807]	PAS_chr3_0951	1,2	Recurrent observation in downstream processes
C4R6G4	Non-ATPase base subunit of the 19S regulatory particle (RP) of the 26S proteasome		ubiquitin-dependent protein catabolic process [GO:0006511]	PAS_chr3_1083	1,2,3	
C4R6G8	Vacuolar aspartyl protease (Proteinase A)	Cathepsin E	aspartic-type endopeptidase activity [GO:0004190]	PAS_chr3_1087	2,3	
C4R6P3	Protein component of the large (60S) ribosomal subunit		structural constituent of ribosome [GO:0003735]	PAS_chr4_0041	2	Recurrent observation in downstream processes
C4R703	Lysophospholipase	Lysosomal Phospholipase A2 (LPLA2)	lysophospholipase activity [GO:0004622]; phosphatidyl phospholipase B activity [GO:0102545]	PAS_chr4_0153	2,3	Leads to polysorbate degradation by cleaving acyl ester bonds of glycerophospholipids
C4R7A1	40S ribosomal protein S4		rRNA binding [GO:0019843]; structural constituent of ribosome [GO:0003735]	PAS_chr4_0246	1,2	Recurrent observation in downstream processes
C4R7W9	Lon protease homolog, mitochondrial		ATP binding [GO:0005524]; ATP hydrolysis activity [GO:0016887]; ATP-dependent peptidase activity [GO:0004176]; serine-type endopeptidase activity [GO:0004252]	PIM1 PAS_chr4_0441	1,2,3	Recurrent observation in downstream processes
C4R7Y4	Protein component of the small (40S) ribosomal subunit		structural constituent of ribosome [GO:0003735]	PAS_chr4_0456	1,2	Recurrent observation in downstream processes
C4R887	ATPase involved in protein folding and nuclear localization signal (NLS)-directed nuclear transport		ATP binding [GO:0005524]; ATP hydrolysis activity [GO:0016887]	PAS_chr4_0552	1,2	
C4R8B8	Aspartic protease, attached to the plasma membrane via a glycosylphosphatidylinositol (GPI) anchor GN=PAS_chr4_0584 PE=3 SV=1		aspartic-type endopeptidase activity [GO:0004190]	PAS_chr4_0584	1,2,3	Aspartic protease
C4R8R1	Cytoplasmic thioredoxin isoenzyme of the thioredoxin system	Periredoxin-1 family	protein disulfide reductase activity [GO:0015035]	PAS_chr4_0725	2	
C4R8U6	Uncharacterized protein GN=PAS_chr4_0989 PE=4 SV=1		GTPase regulator activity [GO:0030695]; metal ion binding [GO:0046872]	PAS_chr4_0989	1,2	
C4R938	Protein disulfide isomerase, multifunctional protein resident in the endoplasmic reticulum lumen	Protein Disulfide Isomerase	Protein disulfide isomerase activity [GO:0003756]	PAS_chr4_0844	2	Affects drug product quality by reduction of protein disulfide bonds between cysteine residues

436

These design criteria were implemented in the synthesis of a One-Bead-One-Peptide (OBOP) combinatorial library on ChemMatrix™ resin, whose translucent, porous, hydrophilic beads are ideal for library selection in competitive mode [27]. To ensure the identification of selective HCP-targeting ligands capable of purifying both full mAbs and engineered mAb-derived products [50] in flow-through mode, we implemented an automated orthogonal fluorescence screening using a bead-imaging-and-sorting device developed in prior work [28]. Our workflow utilizes an automated bead-selection algorithm that processes images of the beads in real time, thus enabling the sampling of a large portion of the library and utilizes multiple fluorescence emission wavelengths to select leads with high target-binding strength and selectivity. To that end, a library screening feedstock mimicking industrial harvests was prepared by combining *K. phaffii* HCPs and human IgG – including whole antibody, Fc/Fab fragments, and nanobody fragments – at a titer of 0.2 mg/mL and 2 mg/mL respectively. The HCPs and IgG species were collectively labeled with a red and a green-fluorescent dye, respectively. The library beads displaying high-intensity red-only fluorescent emission (*i.e.*, strong HCP-only capture behavior at thermodynamic equilibrium) were selected and the peptides carried thereupon were sequenced via Edman degradation, and their sequences were analyzed to identify position-based homology of residues (**Figure S2A**) and amino acid frequency in the identified peptides (**Figure S2B**).

The sequences present a strong enrichment in aromatic and cationic/H-bonding residues, and a depletion in anionic and aliphatic residues. The physicochemical properties of the candidate ligands are complementary to those of *K. phaffii* HCPs, thus providing confidence in the outcome of library screening. We note that, while the enrichment of cationic residues was anticipated, given the abundance of anionic HCPs, the presence of a population of cationic HCPs (**Figure S1B**) seemed to warrant some enrichment of Asp, which was not registered. Accordingly, to ensure the broadest possible HCP-binding activity, eight peptides were selected – namely, RYWV, QEKK, VWHH, EWAK, RYWK, YHKH, RWYQ, WYKK – that feature a diverse amino acid arrangement and composition and are thus expected to encompass a broad spectrum of binding modalities.

3.3. Capture of K. phaffii HCPs via flow-through affinity chromatography using PichiaGuard. The selected peptides were conjugated on Toyopearl beads, a polymethacrylate-based chromatographic resin whose mechanical rigidity, particle size (65 µm), and pore diameter (100 nm) are ideal for protein purification via flow-through affinity chromatography [24]. The *K. phaffii* X-33 cell culture fluid (CCF) utilized to evaluate the HCP binding activity of PichiaGuard was initially diafiltered to *(i)* remove residual glycerol and methanol as well as fragments resulting from proteolytic activity and residual media components, which are likely to interfere with the protein:peptide interaction (**Figure S2C**); *(ii)* adjust the HCP titer to different concentrations, ranging from 0.01 to 1.5 mg/mL to study the effect of protein concentration on binding kinetics and capacity; and *(iii)* adjust the buffer composition to either 20 mM sodium acetate at pH 5.7 (2.6 mS/cm) or 20 mM sodium phosphate at pH 7.1 (7 mS/cm) to study the effect of ionic strength and pH on binding kinetics and capacity. These buffers were adopted owing to their kosmotropic character that ensures a native, folded protein state, while promoting affinity interactions; furthermore, the pH chosen to prepare the acetate buffer matches that of *K. phaffii* culture conditions, while the neutral pH of the phosphate buffer matches that utilized for the CHO HCP-targeting LigaGuard™ and provides a necessary control to interpret the HCP capture results.

Static binding studies were conducted by incubating the PichiaGuard resin with the conditioned harvests and the residual HCP titer in solution were measured to calculate the equilibrium binding. The resulting adsorption points were fit against Langmuir isotherm curves (**Figures 2A** and **2B**), from which the values of maximum binding capacity (Q_{max}) and dissociation constant (K_D) for the peptide mixture were derived. The rather notable difference in Q_{max}, 25.4 and 18.3 mg HCP per mL resin respectively registered in acetate and phosphate buffers, could be attributed to the different ionic strength of the buffers, wherein the higher conductivity of phosphate buffer shields Coulomb interactions; however, the

observation that cationic peptide ligands provide lower binding at pH 7, where the anionic HCPs should be more accentuated, suggests that the electrostatic component of binding is one, but not the dominant, component of the HCP:peptide interaction. Thus, the variation in Q_{max} can also be imputed to the different position in the Hofmeister series of the anions, wherein the less kosmotropic acetate ions promote the formation of ligand-accessible cavities/pockets on HCPs (*i.e.*, salting in) that provide ideal target sites for hydrophobic or aromatic and polar residues in the PichiaGuard peptides to respectively form hydrophobic or π-π and hydrogen bond interactions; at the low concentration in the respective buffers, in fact, the acetate and phosphate ions are not expected to perturb hydrogen bonds and salt bridges between HCPs and peptides [51–53]. The combination of these effects results in a strong HCP-binding activity by the peptide-functionalized resin, and ultimately a high binding capacity.

The values of K_D – 17.5 μM and 30 μM measured in acetate and phosphate buffer, respectively – seem to indicate a moderate HCP:peptide affinity. These values, however, result from treating the *K. phaffii* HCPs as a single species, a characteristic trait of the evaluation of a chromatographic technology against fluids featuring varying HCP titers and profiles, and commonly found in the literature. On the other hand, HCP capture is driven by the values of K_D of the single HCP:peptide binding events, and since each HCP is present at very low concentration (≤ 1 nM), the binding affinity of the selected ligands is actually high, as observed in prior work on LigaGuard™ ligands [54].

(A)
(B)
(C)
(D)

Figure 2: HCP binding isotherms obtained by incubating PichiaGuard TP-650M resin with solutions of K. phaffii HCPs at concentrations ranging between 0 – 1 mg/mL in either *(A)* 20 mM sodium acetate at pH 5.7 or *(B)* 25 mM sodium phosphate at pH 7.2. Values of cumulative logarithmic removal of K. phaffii HCPs afforded by CaptoAdhere, CaptoQ, CHO LigaGuard, and PichiaGuard TP-650M resins loaded (residence time: 1 minute) with X-33 feedstocks with HCP titer of ~0.4 mg/mL and conditioned in either *(C)* 20 mM sodium acetate at pH 5.7 or *(D)* 25 mM sodium phosphate at pH 7.2.

***3.4. Clearance of K. phaffii HCPs in flow-through mode: PichiaGuard vs. ion exchange resins*.** The purification of biopharmaceuticals expressed by *K. phaffii* relies almost ubiquitously on ion exchange (IEX) and mixed-mode (MM) chromatography. In particular, the post-capture steps of product polishing are being increasingly conducted in flow-through mode, which combines speed of operation with lower capital and operational costs. In this context, the IEX and MM resins employed in flow-through operations often include a quaternary amine (Q) moiety, which, as a strong cationic group, provides high binding capacity and strength of HCPs over a broad range of pH and conductivity values [23,55]. Owing to their high capacity, IEX resins are often utilized at the capture step when affinity resins are not available. On the other hand, these resins lack binding selectivity and must be combined in a series of orthogonal chromatographic steps to ensure high product purity, which requires extensive process optimization and is often achieved at the expense of yield [56]. The paradigm of flow-through affinity chromatography as a step dedicated to abating process-related impurities holds strong promise to overcome these challenges and offers an ideal route to de-risk platform processes for established products as well as products that do not benefit from a dedicated affinity resin. In prior work on mAb purification in flow-through mode, we observed that LigaGuard™ resin outperform commercial IEX and MM resins on clarified and unconditioned CHO cell culture harvests, especially due to their lack of needing process optimization [27,54,57,58]. In this study, we resolved to conduct an analogous comparison between PichiaGuard, LigaGuard™ [54], and a commercial anion exchanger (AEX, Capto Q) and cationic MM (CaptoAdhere) resins by tracking HCP capture upon continuous loading of a clarified *K. phaffii* harvest [59].

To study the effect of pore size, the PichiaGuard was conjugated on two different polymethacrylate-based resins, one with pore diameter of 100 nm (Toyopearl 650M, TP-650M) and one with pore diameter of 5 nm (Toyopearl HW-50F, TP50F). Small proteins such as engineered antibody fragments, such as Fab or scFv, feature hydrodynamic radii ~2-4 nm [60,61] and are produced often using *K. phaffii* [62]. Accordingly, we sought to explore the combination of size exclusion of the product and affinity capture of HCPs as a means to improve product yield and purity under optimal conditions of linear velocity of loading [63]. Accordingly, we resolved to operate the flow-through process at bioprocess-relevant values of residence time (RT, 1 minute) to develop an efficient flow-through process. Finally, similar to static binding tests, the harvest was conditioned in either phosphate buffer at pH 7.1 or acetate buffer at pH 5.7. The flow-through effluent was continuously collected and apportioned in fractions at regular intervals to evaluate the temporal profiles of protein binding – both as global HCP titers and single species via proteomics analysis.

The results presented in **Figure 2** provide a comparison of the HCP-clearing performance of the peptide-functionalized resin (PichiaGuard) *vs.* AEX, CHO LigaGuard™ and MM resins as a function of load in the two buffer systems. The PichiaGuard outperformed both LigaGuard™ and commercial resins when operated in acetate buffer at the RT of 1 min, providing a cumulative HCP LRV > 2 (*i.e.*, > 100-fold reduction of HCPs) when loaded with up to 10 CVs, followed by LRVs of 1.5 - 2 (*i.e.*, 30-100-fold reduction) for up to 20 CVs, 1.3 - 1.5 (*i.e.*, 20-30-fold reduction) for up to 40 CVs, and plateauing at 1.3 for up to 100 CVs; LigaGuard™ resin afforded the second best HCP clearance profile, with LRVs of 1 - 1.5 for up to 10 CVs, 0.75 - 1 for up to 20 CVs, and plateaued to 0.45-0.55 CV when loaded up to 40 CVs. The reference AEX and MM resins provided a rather poor HCP clearance (without keen buffer or process optimization), consistently below a 10-fold reduction and plateauing to a LRV ~ 0.5 (3-fold reduction) when loaded up to 40 CVs. A similar performance ranking was observed in phosphate buffer, with the sole exception of the AEX resin, which performed on par with the PichiaGuard resin across the entire loading range (40 CVs). The magnitude of HCP LRVs, however, decreased for all resins, thus mirroring the stating binding results discussed above: the highest drop was registered with PichiaGuard, whose HCP clearance activity in phosphate buffer was 10-fold lower than its acetate counterpart, while the performance of LigaGuard™ and CaptoAdhere was almost unaffected.

A number of conclusions can be drawn from these results: *(i)* peptide ligands significantly outper-formed simple MM or IEX ligands, likely due to their broader range of non-covalent interactions and con-formations; and *(ii)* library selection targeted to *K. phaffii* HCPs proved critical to identify an ensemble of bespoke ligands. While in fact the LigaGuard™ peptides delivered a higher HCP capture than MM or IEX moieties (**Table 2**), likely owed to the variety of interactions peptide ligands offer, the *K. phaffii* HCP-tar-geting peptides perform as custom affinity ligands – or at least, as more advanced than conventional mixed-mode ligands without the need for process optimization – by displaying both stronger binding strength and capacity and host selectivity. Furthermore, *(iii)* the PichiaGuard peptides ought to be utilized under conditions, such as the 20 mM acetate buffer at pH 5.7, that resemble the native physicochemical environment of *K. phaffii* harvests, namely acidic pH. Finally, *(iv)* around the 30th CV of loading, we ob-served the cumulative LRV plateau at ~1 and the corresponding fractional LRV become lower than 0.5 (**Figure S3E**), indicating that the adsorbent has reached saturation. The 30th CV mark corresponds to a load ratio of ~18 g of HCPs per L of resin, which corresponds to ~75% of Q_{max} (see *Section 3.3*). This is a reason-able outcome given the relationship between dynamic *vs.* static binding and can be improved by adjusting the residence time as well as buffer composition and pH. Nonetheless, contrary to affinity resins that operate in bind-and-elute mode, the loading of 'Guard' resins is intended to reach halfway between the values of $DBC_{10\%}$ and Q_{max}.

Table 2. *Number of K. phaffi HR-HCPs captured by PichiaGuard, CHO LigaGuard™, and CaptoQ resins loaded with X-33 feedstocks (residence time: 1 minute) with HCP titer of ~0.4 mg/mL and conditioned in either 20 mM sodium acetate at pH 5.7- or 25-mM sodium phosphate at pH 7.2.*

Acetate				
	CaptoQ	CHO LigaGuard™	PichiaGuard-TP50F	PichiaGuard-TP650M
Undetected	19	18	19	18
Not Captured	15	19	5	10
Captured	10	11	20	16
%	40%	42%	80%	62%
Phosphate				
	CaptoQ	CHO LigaGuard™	PichiaGuard-TP50F	PichiaGuard-TP650M
Undetected	0	8	0	0
Not Captured	29	20	17	8
Captured	15	15	27	36
%	34%	41%	61%	82%

Improving the clearance of HCPs from a given *K. phaffii* harvest using PichiaGuard resin can be easily achieved via process optimization. Conversely, achieving comparable purification under given conditions from fluids that differ substantially in the profile of physicochemical properties of their HCPs is far from a foregone conclusion. We therefore resolved to measure the global HCP removal from two fluids – pro-duced respectively with and without the addition of protease inhibitors to the cell culture medium – whose distribution of HCP molecular weights differed radically (**Figures 2C** and **2D**). To that end, we im-plemented a design-of-experiments (DOE) algorithm to explore systematically the effects of both categor-ical (*i.e.*, buffer and resin) and continuous (*i.e.*, buffer conductivity and pH, load ratio, and RT) variables to identify parameters that maximize HCP capture (**Figure S3A-D**). The choice of resin, buffer, and interaction effects between these parameters were indeed found to be significant. The effect of RT appeared to be the most pronounced when using phosphate buffer: the highest HCP reduction (LRV ~3) was obtained at the RT of 1 min, which was therefore adopted for the remainder of this study. Notably, the combination of phosphate buffer and a smaller pore diameter of the resin led to higher HCP clearance (LRV ~ 0.8 - 1)

than its acetate counterpart (LRV ~ 0.7 – 0.9), possibly due to effects of phosphate counterions on protein folding and increased ability to diffuse into smaller pores; however, limiting the total capture capacity. PichiaGuard-TP650M, however, afforded a superlative HCP reduction (LRV ~ 3) and emerged as the resin of choice. As observed above, performing flow-through purification in acetate buffer consistently achieved a higher HCP clearance thus providing conclusive evidence for the adoption of acetate buffer for resin equilibration and, ideally, conditioning of the feedstock.

3.5. Tracking the removal of high-risk HCPs from K. phaffii cell culture harvests by PichiaGuard. In the established biopharmaceutical practice, the validation of a batch of therapeutic proteins consists in certifying that the residual HCP and hcDNA titers are below an acceptable limit for monoclonal antibodies (mAbs), these values are 100 ppm and 10 ppb, respectively [64]. The measurement of residual HCP titer has been conducted for decades with ELISA kits. Like all assays that relies on polyclonal antibodies raised against multiple antigens, most ELISAs do not provide a full coverage of HCPs, but only of those that are abundant or more immunogenic to the animal host; furthermore, HCP aggregation and the formation of HCP:product complexes – a widespread phenomenon in bioprocess fluids – can impair assay readout [29]. These elements, combined with the variability among assay lots and operator's performance, question the accuracy of HCP quantification provided by ELISA assays. Furthermore, in recent years, the growth of proteomics in biomanufacturing has shown that product batches with acceptable global level of impurities can contain amounts of single high-risk HCPs (HR-HCPs) that pose a serious threat to patient health (*e.g.*, are toxic or immunogenic) or can degrade the mAb or its excipients during storage, resulting in reduced efficacy or harmful products [29,65]. Recent studies have amply documented that commercial chromatographic resins may struggle to remove particular HR-HCPs: a number of these "persistent" HR-HCPs have been reported on both a process- and product-basis and have been linked to delays in clinical trials and process approval or the recall of drug batches.

Under this premise, we resolved to complement the measurements of global residual HCPs presented above with tracking the individual HCPs captured by PichiaGuard via proteomics analysis. To this end, the effluents obtained in both acetate and phosphate buffers at the RT of 1 min were analyzed via LC-MS/MS by implementing the proteomics workflow outlined above for top-speed data dependent acquisition (DDA) [66]. Spectral interrogation using the Proteome Discoverer Suite was used to perform label-free quantification (LFQ) of HCPs in the flow-through fractions *vs.* the corresponding feedstocks (*note:* due to the lower number of proteins identified in *K. phaffii* harvests compared to HCP harvests, a 'match between runs' was implemented to preserve information from low-intensity peptide hits that that may have been masked by abundant contaminants like trypsin and any media components) [67]. Briefly, *(i)* data processing prior to analysis of variance (ANOVA) involved the exclusion of contaminants such as trypsin and keratin as well as the HCPs found in the effluents but not the load sample; *(ii)* the abundance of individual HCPs was used to calculate the removal values [68]. We noted that certain proteins were identified in some flow-through fractions to have an abundance ratio > 1 (defined as the ratio of HCP abundance in the effluent sample *vs.* the load sample), suggesting that their presence in the load was masked by others in the load sample, more abundant peptides with similar sequence or charge which were not excluded.

We adopted three main indicators to evaluate the HCP-capture performance of different adsorbents, namely *(a)* global concentration and reduction; *(b)* between-group analysis, which presents the abundance trends of every HCP identified; and *(c)* clearance or persistence of high-risk HCPs (HR-HCPs). These indicators capture different variations within our study group and have been used collectively to make inferences. **Figure 3** presents the global landscape of the HCPs identified across all control experiments. The between-group (group = experiment of a particular resin with a particular buffer system) significance analysis, estimated using an iterative Tukey-Kramer HSD test known as the Newman-Kuels method between all pairs was used to construct a 'connected-color' plot, which connects fractions with a similar normalized mean abundance with the same color and ensues discontinuity with significant 'within group'

difference (p-value > 0.01). This test provided a picture of changes in overall abundance compared to the load and other fractions **(Table S3)** while other statistics and visualization techniques have been employed to track changes across individual species.

To visualize and compare HCPs in successive flow-through fractions within and across buffer groups – (acetate, phosphate) and adsorbents (PichiaGuard-TP650M, PichiaGuard-TP50F resins, CHO LigaGuard™ and Capto Q resin) – **Figure 3B** presents a reduced-dimension view of individual HCP abundance (Uniprot Accession numbers, left-to-right) *vs.* volume of effluent (flow-through fraction, bottom-to-top). Specifically, abundance ratios (fraction by load) are represented as contour plots wherein ratios ≥ 1 (in red) mark HCPs that were not cleared or whose presence was masked in the feedstock by other species but was detected in a flow-through fraction, whereas ratios ≤ 1 (veering from yellow to green as the value decreases) mark HCPs that were captured. At a glance, we observed the ability of PichiaGuard resins to associate with most HCPs within both buffer systems compared to the reference resin counterparts. The contour plots provide visual indication of individual HCPs that were not captured throughout a significant part of the load due to insufficient HCP:ligand binding strength or possibly due to lack of charge/polarity of HCPs at the pH condition of the experiment, based on their isoelectric points.

Notably, the proteomics analysis of the effluents allowed us to track the capture of *K. phaffii* HR-HCPs by the various adsorbents (**Figure 3C**), specifically focusing on proteins that are *(i)* homologous to known high-risk and persistent CHO HCPs; *(ii)* immunogenic, either reported or predicted based on their ability to generate MHC-II binding peptides; and *(iii)* possess proteolytic or other enzymatic activity, as outlined in *Section 3.1*. The number of bound HR-HCPs across groups were classified according to their risk subtypes/bins (**Figure 3D**). Corresponding logarithmic values of abundance variation of all identified HR-HCPs is presented in **Figures S4A** and **S4B**, whereas the number of captured HR-HCPs as a function of loaded volume across all groups is reported in **Figure S4C**.

From these results, the following key observations were drawn. Firstly, the PichiaGuard resins (**Figure 3B**, bottom row) outperformed both LigaGuard™ and AEX (Capto Q) resins in retaining a higher number of *K. phaffii* HCPs, thus corroborating the results of total HCP quantification obtained from the total protein assays (see *Section 3.5*). It is noteworthy that this behavior is observed across all flow-through fractions for the HCPs, indicating a true lack of capture. As shown in **Figure S4C**, of the 187 HCPs identified in the acetate-conditioned feedstock, 180 were captured by PichiaGuard resins throughout the entire feedstock loading, while 3 – 5 were not captured; similarly, of the 318 HCPs in phosphate-conditioned feedstock (for a total of 335 unique *K. phaffii* HCPs identified), 302 were captured, while 8 – 12 were not captured. Secondly, PichiaGuard-TP650M and PichiaGuard-TP50F resin performed similarly in acetate buffer, whereas in phosphate buffer PichiaGuard-TP650M clearly outperformed its TP50F counterpart. The contour plots also show a greater extent of HCP dissociation from PichiaGuard resins in acetate media - especially from TP650M resins, likely a combined effect of pore size and residence time. Thirdly, PichiaGuard demonstrated the ability to clear up to 82% of HR-HCPs – including various aspartic proteases, ion protease homologs, etc. – whereas LigaGuard™ and CaptoQ resins only bound up to 42% of them. This corroborates our claim that ion-exchange resins, despite their satisfactory global HCP capture under optimized conditions [69], can fail to clear a number of HR-HCPs and are therefore unlikely to serve effectively as an HCP-scrubbing step prior to or in lieu of the affinity-based product capture step. Conversely, PichiaGuard makes an excellent tool for orthogonal HCP removal, thus safeguarding the performance and lifetime of affinity resins as well as promoting product quality and stability.

As is to be expected with any chromatographic adsorbent, process optimization in terms of loading ratio, flow rate, and buffer composition is to some extent needed. In the context of this study, however, these results aim to demonstrate the potential of PichiaGuard resin to increase the performance and robustness of current processes as well as provide a route towards platform processes with altogether novel design for isolating biopharmaceutics from *Pichia* harvests.
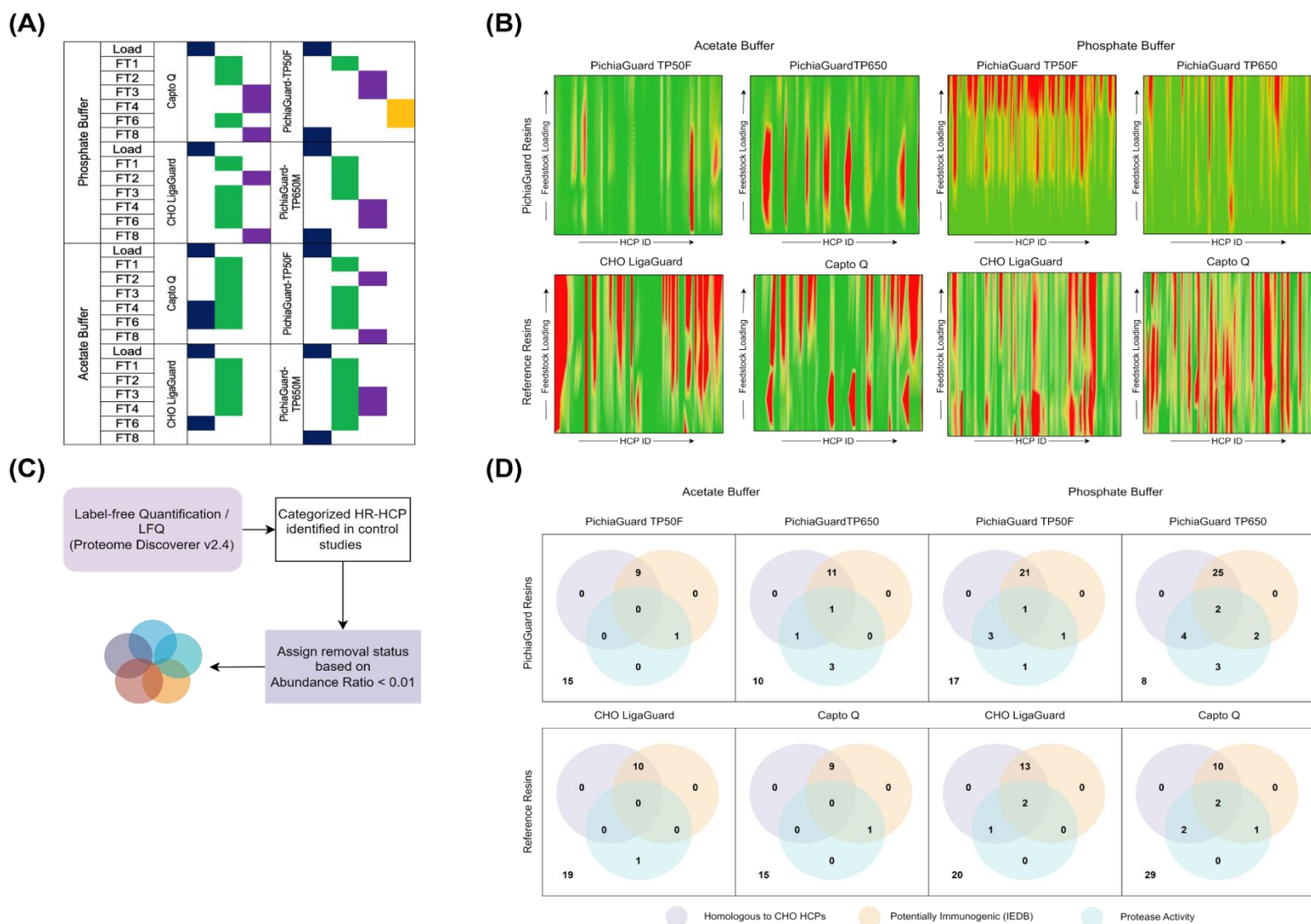
(A)

(B)

(C)

(D)

Figure 3: (A) Connected colors plot indicating significant differences in mean HCP abundance among flow-through fractions; groups connected with different colors feature significantly different HCP abundances calculated as an average value for the group; (B) Contour plots of K. phaffii HCP abundance ratios registered in the effluents obtained by loading CaptoQ, LigaGuard, and PichiaGuard resins loaded (residence time: 1 minute) with X-33 feedstocks with HCP titer of ~0.4 mg/mL and conditioned in either 20 mM sodium acetate at pH 5.7 or 25 mM sodium phosphate at pH 7.2: individual HCPs are aligned on the x-axis, while the number of collected flow-through fractions is on the y-axis; red to green indicates high to low ratio of individual HCP in the effluent vs. feedstock as detected by nanoLC-MS; (C) Flowchart describing the designation of removal status for 'high-risk' HCP (HR-HCPs); (D) Number of individual HR-HCPs removed by CaptoQ, LigaGuard, and PichiaGuard resins grouped by risk category; the values outside the Venn diagrams indicate the number of HCPs that were not removed.

3.6. Purification of ScFv13R4 and mAb from K. phaffii harvests via flow-through affinity chromatography using PichiaGuard. Purification processes fully operated in flow-through mode are the epitome of continuous downstream biomanufacturing: their design minimizes process footprint and the number of tanks for storing service buffers, simplifies the pipework and control systems, and accelerates operations thus increasing productivity. Straight-through processes combining column chromatography and filtration unit operations have been proposed for the continuous manufacturing of mAbs from CHO cell harvests [20,59]. These processes can be successfully designed based on the size of the target product, which is consistently larger than most of the process-related impurities. Conversely, small proteins are significantly more challenging products (**Table S1**), as they remove filtration as an orthogonal method of separation and force reliance solely on chromatography. In this context, pseudo-continuous processes can be utilized by integrating periodic counter-current (PCC) or simulated moving bed (SMB) modes [70] encompassing ion exchange chromatography, whose complexity increases both capital and operation costs. An effective alternative is offered by leveraging resins specifically tasked with HCPs and hcDNA clearance, like LigaGuard™ and PichiaGuard, which provide a product-agnostic route towards fully chromatographic straight-through processes for the purification of both mAb and non-mAb drugs.

Under this premise, having demonstrated the broad HCP-capture activity of PichiaGuard, we moved to evaluate its ability to purify therapeutic proteins from *K. phaffii* harvests in flow-through mode. To this end, we adopted an antibody fragment (ScFv13R4, MW ~30 kDa and pI of 8.36) and a full antibody (referred to as 'mAb' hereon, MW ~150 kDa and pI of 7.56) as model targets. Most importantly, to demonstrate the robustness of the flow-through affinity technology, the *K. phaffii* harvests were loaded on PichiaGuard post diafiltration into acetate buffer, until reaching a ratio >15 grams of HCPs per liter of resin (corresponding to $DBC_{20\%}$, *Section 3.4*). The analyses of the flow-through fractions – namely, global clearance of HCP and other impurities estimated and product yield – are reported in **Table 3**.

Table 3. *Results of ScFv13R4 purification obtained by loading PichiaGuard-TP650 resin loaded (residence time: 1 minute) with X-33 K. phaffi feedstock with ScFv13R4 titer of ~ 1 mg/mL, HCP titer of ~0.4 mg/mL, and conditioned in 20 mM sodium acetate at pH 5.7.*

Molecule	Protein Loading (g/L resin)	Ratio of product:non-product peaks (RP-HPLC)	Cumulative HCP LRV	Cumulative DNA LRV	Cumulative ScFv Yield
	Load	0.19	- - -		
ScFv13R4	2.10	< LOD	3.00	1.23	54.16%
	3.90	< LOD	2.55		
	5.10	2.06	1.92		
	6.30	2.83	1.55		
	7.50	1.77	1.34		
	9.30	1.40	1.14		
	14.40	1.09	1.01		

The purification of ScFv13R4 returned considerable values of impurity removal, with HCP clearance ranging from LRV > 2.5 (~320-fold removal) to LRV > 1.5 (~32-fold removal) when loading respectively up to 3.9 and 7.5 grams of HCPs per liter of resin, finally reaching LRV ~ 1 at loading 14.4 mg/mL resin, at which a DNA LRV ~1.25 was also achieved. The electrophoretic analysis of the flow-through fractions in **Figure 4A** provides at-a-glance presentation of impurity removal by PichiaGuard: particularly noteworthy is the comparison between the feedstock and effluents containing the ScFv13R4, which demonstrate the removal of contaminants both heavier and lighter than the product. Notably, the proteomics analysis of the effluents showed that, of the total 336 HCPs identified in the feedstock, the number of escaping species increased from 11 at the initial collection point of 4 g/L to merely 34 at the final pool, with 10 not

being captured throughout. This corresponds to > 84% of HCPs completely removed – chiefly among them, 38 out of 44 HR-HCPs – while the remainder ~15% HCPs were partially cleared. At the same time, the yield of ScFv13R4 reached only ~56% at the end of the loading phase (estimated using 660 nm BCA for 6x-His tagged ScFv13R4) but was recovered from the column by 'chasing' it with a high-conductivity buffer (**Figure S5B,** chase not quantified). While the low titer of ScFv13R4 in the feed (~1 mg/L) made product detection challenging, the analysis of the flow-through fractions collected at a load value of > 14 mg/mL resin demonstrated a level of HCP-HCP or HCP-product association (**Table 3** and **Figure S5C**); specifically, the reverse phase chromatograms show that the product-to-impurity ratio in the effluent increases as the loading progresses, translating in a growing product enrichment throughout the flow-through operation. The relationship between HCP-capture using PichiaGuard resins as a function of ScFv13R4 product concentration will be studied as the next step. Similar results were obtained from the purification of a full mAb (**Figure S5A**): overall HCP clearance of 1.15 LRV was obtained up to a load of 15 mg/mL with a mAb product recovery of 80% estimated using an analytical Protein A HPLC assay.

Collectively, these results provide a strong proof-of-concept of PichiaGuard as a product-agnostic technology for impurity removal from *K. phaffii* cultures. At the same time, improvement in binding selectivity via optimizing ligand combinations or pretreatment of these cultures needs to be explored, along with the development of additional analytics.
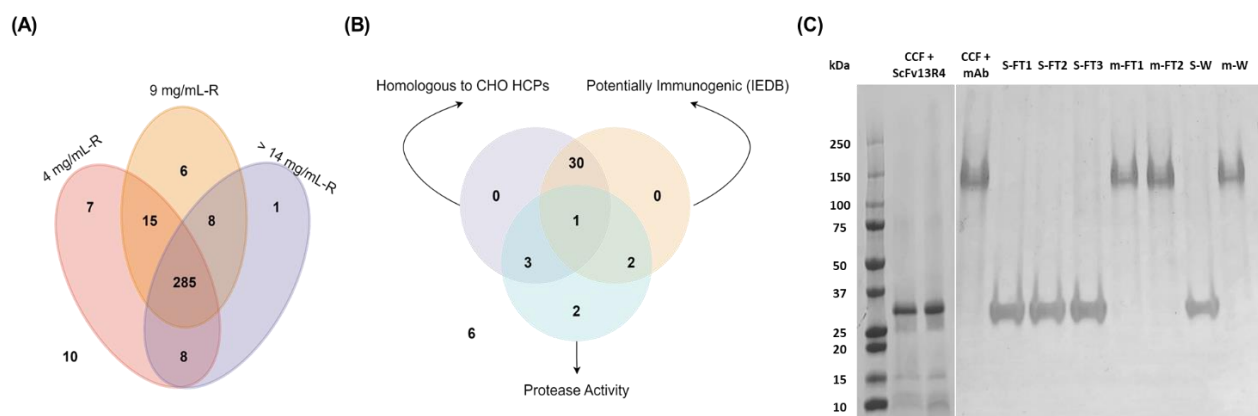

Figure 4: (A) *Total number of HCPs removed from a X-33 K. phaffi feedstock (ScFv13R4 titer of ~ 1 mg/mL, HCP titer of ~0.4 mg/mL, 20 mM sodium acetate at pH 5.7) by PichiaGuard-TP650 resin operated at the residence time of 1 minute as a function of resin loading (namely, 4, 9, and >14 mg of proteins per mL of resin); the number of removed HCPs was determined via proteomic analysis of the flow-through fractions via nanoLC-MS/MS; (B) Number of individual HR-HCPs removed by PichiaGuard-TP650 resin grouped by risk category; the values outside the Venn diagrams indicate the number of HCPs that were not removed; (C) SDS-PAGE gel (native conditions) of the K. phaffi cell culture harvest containing ScFv13R4 and mAb (lanes 2 – 4), and the flow-through fractions generated by loading the harvests on PichiaGuard-TP650 resin at the residence time of 1 minute (5 – 7 for ScFv13R4; 8 and 9 for mAb) and final column wash (lane 8 for scFv13R4 and lane 9 for mAb).*

4. Discussion and Conclusions. The success of *K. phaffii* in pharmaceutical biomanufacturing is scripted in its rapid growth in chemically defined media, low susceptibility to virus contamination [71], facile expression inducibility, and ability to secrete human proteins with correct post-translational modifications at high titer and purity [8][72]. These upstream-related benefits call for convergent efforts in the downstream toolbox, especially in the context of continuous processing and isolation of products for which dedicated affinity adsorbents are not available. In this context, new chromatographic tools are needed that provide orthogonality in removing process-related impurities and/or complementary to commercial adsorbents in clearing residual impurities to ensure the safety of patients treated with complex biological drugs [48,73]. Our group has contributed to this field by introducing the paradigm of flow-through affinity

chromatography in the form of 'Guard' resins, whose HCP and hcDNA capture activity provides an orthogonal step to affinity resins for product capture or a complementary step to IEX and MM resins for product polishing [54,57]. In this context, we leveraged the heuristic power of designed peptide libraries coupled with high-throughput dual-fluorescence screening to identify peptide ligands that capture persistent and high-risk contaminants; furthermore, we integrated advanced analytics, such as proteomics, to demonstrate the potential of these ligands in biomanufacturing. Our initial efforts, conducted in the domain of full mAb and CHO HCPs, were supported by an established portfolio of analytical technologies and a wealth of literature highlighting target HCPs.

In this study, we extended the 'Guard' technology to *K. phaffii* HCPs by introducing PichiaGuard, the first affinity adsorbent dedicated to the clearance of process-related impurities present in *Pichia* cell culture harvests. While the selection of *K. phaffii* HCP-binding peptides was streamlined by a robust ligand identification technology, the evaluation of the resulting PichiaGuard adsorbent faced challenges related to the quantification of product recovery and contaminant removal. In particular, the dearth of relevant literature on the toxicological, immunological, bioprocess-relevant aspects of *Pichia* HCPs compelled us to formulate criteria for identifying species that pose – or are expected to pose – a risk to patient's health (*e.g.*, immunogenicity via B-cell activation [74]) or product stability, and can compromise the efficiency of the purification process (*e.g.*, co-elution with the product or persistence throughout the downstream train [33]). Having established an analytical panel, we demonstrated that PichiaGuard affords up to 99% reduction of HCPs and hcDNA from a native *K. phaffii* cell culture supernatant, when challenged with a harvest containing a therapeutic product – a full mAb or an scFv fragment – thus providing product enrichment just by just impurity removal. Most importantly, PichiaGuard's ability to clear immunogenic (*e.g.*, 40S ribosomal proteins, Histone H3, Thioredoxin peroxidase, Phosphatidylinositol transfer protein, Translation initiation factor eIF4G etc.), proteolytically (*e.g.*, Lon protease, Aspartic proteases – vacuolar, GPI-anchored and plasma membrane-attached forms) and enzymatically active (*e.g.*, Lysophospholipase, Catalase, Peptidase A1, Ydj1p cooperating heat shock protein etc.) high-risk HCPs is key to process robustness and product safety.

These results warrant a path forward for PichiaGuard, chiefly its integration with commercial IEX and MM resins in a straight-through process. Accordingly, future studies on PichiaGuard will focus on flexibility, whether it can be utilized in a fully product-agnostic manner; robustness, whether it can sustain variability in HCP titer and complexity as well as conductivity and pH of the feedstock; yield, especially focusing on the recovery of HCP-bound product via 'chasing' wash; lifetime, whether it can be regenerated and re-used safely over a high number of cycles. Setting the stage for such work, this study presents a cogent message that the PichiaGuard technology holds true promise to de-risk biomanufacturing, reduce the cost and time-to-market of established therapeutics, and bring to fruition emerging medicines (*e.g.*, cytokines, enzymes, engineered antibodies, and viral vaccines) [1,75][76]. In so doing, the Guard peptide ligands are not anticipated to impose major financial burdens, being manufactured affordably at large scale, or regulatory concerns, given their safety and ease of clearance (small size). This strengthens our confidence in their adaptability in next-generation, truly continuous processes that can affordably and reliably meet the growing global demand [20,77,78].

Conflicts of Interest. The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

[1]   Harzevili, F. D. *Synthetic Biology of Yeasts*; 2022.

[2]   Baghban, R.; Farajnia, S.; Rajabibazl, M.; Ghasemi, Y.; Mafi, A. A.; Hoseinpoor, R.; Rahbarnia, L.; Aria, M. Yeast Expression Systems: Overview and Recent Advances. *Mol. Biotechnol.* 2019, *61*, 365.

[3]   Vuree, S. Chapter 17 - Pichia Pastoris Expression System: An Impending Candidate to Express Protein in Industrial and Biopharmaceutical Domains; Singh, J., Gehlot, P. B. T.-N. and F. D. in M. B. and B., Eds.; Elsevier, 2020; pp 223.

[4]   Huang, C. J.; Damasceno, L. M.; Anderson, K. A.; Zhang, S.; Old, L. J.; Batt, C. A. A Proteomic Analysis of the Pichia Pastoris Secretome in Methanol-Induced Cultures. *Appl. Microbiol. Biotechnol.* 2011, *90*, 235.

[5]   Hou, R.; Gao, L.; Liu, J.; Liang, Z.; Zhou, Y. J.; Zhang, L.; zhang, Y. Comparative Proteomics Analysis of Pichia Pastoris Cultivating in Glucose and Methanol. *Synth. Syst. Biotechnol.* 2022, *7*, 862.

[6]   Lin, X. Q.; Liang, S. L.; Han, S. Y.; Zheng, S. P.; Ye, Y. R.; Lin, Y. Quantitative ITRAQ LC-MS/MS Proteomics Reveals the Cellular Response to Heterologous Protein Overexpression and the Regulation of HAC1 in Pichia Pastoris. *J. Proteomics* 2013, *91*, 58.

[7]   Zahrl, R. J.; Peña, D. A.; Mattanovich, D.; Gasser, B. Systems Biotechnology for Protein Production in Pichia Pastoris. *FEMS Yeast Res.* 2017, *17*, 1.

[8]   Nieto-Taype, M. A.; Garcia-Ortega, X.; Albiol, J.; Montesinos-Seguí, J. L.; Valero, F. Continuous Cultivation as a Tool Toward the Rational Bioprocess Development With Pichia Pastoris Cell Factory. *Front. Bioeng. Biotechnol.* 2020, *8*, 1.

[9]   Brady, J. R.; Whittaker, C. A.; Tan, M. C.; Kristensen 2nd, D. L.; Ma, D.; Dalvie, N. C.; Love, K. R.; Love, J. C. Comparative Genome-Scale Analysis of Pichia Pastoris Variants Informs Selection of an Optimal Base Strain. *Biotechnol. Bioeng.* 2020, *117*, 543.

[10]   Love, K. R.; Shah, K. A.; Whittaker, C. A.; Wu, J.; Bartlett, M. C.; Ma, D.; Leeson, R. L.; Priest, M.; Borowsky, J.; Young, S. K.; Love, J. C. Comparative Genomics and Transcriptomics of Pichia Pastoris. *BMC Genomics* 2016, *17*, 1.

[11]   Mattanovich, D.; Graf, A.; Stadlmann, J.; Dragosits, M.; Redl, A.; Maurer, M.; Kleinheinz, M.; Sauer, M.; Altmann, F.; Gasser, B. Genome, Secretome and Glucose Transport Highlight Unique Features of the Protein Production Host Pichia Pastoris. *Microb. Cell Fact.* 2009, *8*, 29.

[12]   Peg, B.; Pharmaceuticals, E.; Valli, M.; Grillitsch, K.; Grünwald-Gruber, C.; Tatto, N. E.; Hrobath, B.; Klug, L.; Ivashov, V.; Hauzmayer, S.; Koller, M.; Tir, N.; Leisch, F.; Gasser, B.; Graf, A. B.; Altmann, F.; Daum, G.; Mattanovich, D. A Subcellular Proteome Atlas of the Yeast Komagataella Phaffii. *Bioconjugate Tech.* 2013, *20*, xv.

[13]   Love, K. R.; Politano, T. J.; Panagiotou, V.; Jiang, B.; Stadheim, T. A.; Love, J. C. Systematic Single-Cell Analysis of Pichia Pastoris Reveals Secretory Capacity Limits Productivity. *PLoS One* 2012, *7*, 1.

[14]   Staudacher, J.; Rebnegger, C.; Dohnal, T.; Landes, N.; Mattanovich, D.; Gasser, B. Going beyond the Limit: Increasing Global Translation Activity Leads to Increased Productivity of Recombinant Secreted Proteins in Pichia Pastoris. *Metab. Eng.* 2022, *70*, 181.

[15]   Fischer, J. E.; Glieder, A. Current Advances in Engineering Tools for Pichia Pastoris. *Curr. Opin. Biotechnol.* 2019, *59*, 175.

[16]   Juturu, V.; Wu, J. C. Heterologous Protein Expression in Pichia Pastoris: Latest Research Progress and Applications. *ChemBioChem* 2018, *19*, 7.

[17] Ahmad, M.; Winkler, C. M.; Kolmbauer, M.; Pichler, H.; Schwab, H.; Emmerstorfer-Augustin, A. Pichia Pastoris Protease-Deficient and Auxotrophic Strains Generated by a Novel, User-Friendly Vector Toolbox for Gene Deletion. *Yeast* 2019, *36*, 557.

[18] Zhang, Y.; Liu, R.; Wu, X. The Proteolytic Systems and Heterologous Proteins Degradation in the Methylotrophic Yeast Pichia Pastoris. *Ann. Microbiol.* 2007, *57*, 553.

[19] Molden, R.; Hu, M.; Yen E, S.; Saggese, D.; Reilly, J.; Mattila, J.; Qiu, H.; Chen, G.; Bak, H.; Li, N. Host Cell Protein Profiling of Commercial Therapeutic Protein Drugs as a Benchmark for Monoclonal Antibody-Based Therapeutic Protein Development. *MAbs* 2021, *13*.

[20] Crowell, L. E.; Lu, A. E.; Love, K. R.; Stockdale, A.; Timmick, S. M.; Wu, D.; Wang, Y. A.; Doherty, W.; Bonnyman, A.; Vecchiarello, N.; Goodwine, C.; Bradbury, L.; Brady, J. R.; Clark, J. J.; Colant, N. A.; Cvetkovic, A.; Dalvie, N. C.; Liu, D.; Liu, Y.; Mascarenhas, C. A.; Matthews, C. B.; Mozdzierz, N. J.; Shah, K. A.; Wu, S. L.; Hancock, W. S.; Braatz, R. D.; Cramer, S. M.; Love, J. C. On-Demand Manufacturing of Clinical-Quality Biopharmaceuticals. *Nat. Biotechnol.* 2018, *36*, 988.

[21] De Groot, A. S. ISPRI - EpiVax.

[22] Duke, B. R.; Mitra-Kaushik, S. Current In Vitro Assays for Prediction of T Cell Mediated Immunogenicity of Biotherapeutics and Manufacturing Impurities. *J. Pharm. Innov.* 2019, *15*, 202.

[23] Heiss, S.; Maurer, M.; Hahn, R.; Mattanovich, D.; Gasser, B. Identification and Deletion of the Major Secreted Protein of Pichia Pastoris. *Appl. Microbiol. Biotechnol.* 2013, *97*, 1241.

[24] Sripada, S. A.; Chu, W.; Williams, T. I.; Teten, M. A.; Mosley, B. J.; Carbonell, R. G.; Lenhoff, A. M.; Cramer, S. M.; Bill, J.; Yigzaw, Y.; Roush, D. J.; Menegatti, S. Towards Continuous MAb Purification: Clearance of Host Cell Proteins from CHO Cell Culture Harvests via "Flow-through Affinity Chromatography" Using Peptide-Based Adsorbents. *Biotechnol. Bioeng.* 2022, *119*, 1873.

[25] Wiśniewski, J. R. Filter Aided Sample Preparation – A Tutorial. *Anal. Chim. Acta* 2019, *1090*, 23.

[26] Martineau, P.; Jones, P.; Winter, G. Expression of an Antibody Fragment at High Levels in the Bacterial Cytoplasm11Edited by J. Karn. *J. Mol. Biol.* 1998, *280*, 117.

[27] Lavoie, R. A.; Fazio, A.; Williams, T. I.; Carbonell, R.; Menegatti, S. Targeted Capture of Chinese Hamster Ovary Host Cell Proteins: Peptide Ligand Binding by Proteomic Analysis. *Biotechnol. Bioeng.* 2020, *117*, 438.

[28] Saberi-Bosari, S.; Omary, M.; Lavoie, A.; Prodromou, R.; Day, K.; Menegatti, S.; San-Miguel, A. Affordable Microfluidic Bead-Sorting Platform for Automated Selection of Porous Particles Functionalized with Bioactive Compounds. *Sci. Rep.* 2019, *9*, 7210.

[29] Pilely, K.; Johansen, M. R.; Lund, R. R.; Kofoed, T.; Jørgensen, T. K.; Skriver, L.; Mørtz, E. Monitoring Process-Related Impurities in Biologics – Host Cell Protein Analysis. 2022, 747.

[30] Huang, Y.; Molden, R.; Hu, M.; Qiu, H.; Li, N. Toward Unbiased Identification and Comparative Quantification of Host Cell Protein Impurities by Automated Iterative LC–MS/MS (HCP-AIMS) for Therapeutic Protein Development. *J. Pharm. Biomed. Anal.* 2021, *200*, 114069.

[31] Zhang, Q.; Goetze, A. M.; Cui, H.; Wylie, J.; Tillotson, B.; Hewig, A.; Hall, M. P.; Flynn, G. C. Characterization of the Co-Elution of Host Cell Proteins with Monoclonal Antibodies during Protein A Purification. *Biotechnol. Prog.* 2016, *32*, 708.

[32] Aboulaich, N.; Chung, W. K.; Thompson, J. H.; Larkin, C.; Robbins, D.; Zhu, M. A Novel Approach to Monitor Clearance of Host Cell Proteins Associated with Monoclonal Antibodies. *Biotechnol. Prog.* 2014, *30*, 1114.

[33] Jones, M.; Palackal, N.; Wang, F.; Gaza-Bulseco, G.; Hurkmans, K.; Zhao, Y.; Chitikila, C.; Clavier, S.; Liu, S.; Menesale, E.; Schonenbach, N. S.; Sharma, S.; Valax, P.; Waerner, T.; Zhang, L.;

Connolly, T. "High-risk" Host Cell Proteins (HCPs): A Multi-company Collaborative View. *Biotechnol. Bioeng.* 2021.

[34] Hammond, S.; Kaplarevic, M.; Borth, N.; Betenbaugh, M. J.; Lee, K. H. Chinese Hamster Genome Database: An Online Resource for the CHO Community at Www.CHOgenome.Org. *Biotechnol. Bioeng.* 2012, *109*, 1353.

[35] Bailey-Kellogg, C.; Gutiérrez, A. H.; Moise, L.; Terry, F.; Martin, W. D.; De Groot, A. S. CHOPPI: A Web Tool for the Analysis of Immunogenicity Risk from Host Cell Proteins in CHO-Based Protein Production. *Biotechnol. Bioeng.* 2014, *111*, 2170.

[36] Zahrl, R. J.; Peña, D. A.; Mattanovich, D.; Gasser, B. Systems Biotechnology for Protein Production in Pichia Pastoris. *FEMS Yeast Res.* 2017, *17*, 68.

[37] Renuse, S.; Madugundu, A. K.; Kumar, P.; Nair, B. G.; Gowda, H.; Prasad, T. S. K.; Pandey, A. Proteomic Analysis and Genome Annotation of Pichia Pastoris, a Recombinant Protein Expression Host. *Proteomics* 2014, *14*, 2769.

[38] Pichia Genome Repository www.pichiagenome.org.

[39] Baumann, K.; Carnicer, M.; Dragosits, M.; Graf, A. B.; Stadlmann, J.; Jouhten, P.; Maaheimo, H.; Gasser, B.; Albiol, J.; Mattanovich, D.; Ferrer, P. A Multi-Level Study of Recombinant Pichia Pastoris in Different Oxygen Conditions. *BMC Syst. Biol.* 2010, *4*, 141.

[40] Brady, J. R.; Whittaker, C. A.; Tan, M. C.; Kristensen, D. L.; Ma, D.; Dalvie, N. C.; Love, K. R.; Love, J. C. Comparative Genome-Scale Analysis of Pichia Pastoris Variants Informs Selection of an Optimal Base Strain. *Biotechnology and Bioengineering*. 2020, pp 543.

[41] Naranjo, C. A.; Jivan, A. D.; Vo, M. N.; de Sa Campos, K. H.; Deyarmin, J. S.; Hekman, R. M.; Uribe, C.; Hang, A.; Her, K.; Fong, M. M.; Choi, J. J.; Chou, C.; Rabara, T. R.; Myers, G.; Moua, P.; Thor, D.; Risser, D. D.; Vierra, C. A.; Franz, A. H.; Lin-Cereghino, J.; Lin-Cereghino, G. P. Role of BGS13 in the Secretory Mechanism of Pichia Pastoris. *Appl. Environ. Microbiol.* 2019, *85*.

[42] Lavoie, R. A.; Williams, T. I.; Blackburn, R. K.; Carbonell, R. G.; Menegatti, S. Development of Peptide Ligands for Targeted Capture of Host Cell Proteins from Cell Culture Production Harvests. In *Methods in molecular biology (Clifton, N.J.)*; United States, 2021; Vol. 2261, pp 489.

[43] Hsiao, N. W.; Chen, Y.; Kuan, Y. C.; Lee, Y. C.; Lee, S. K.; Chan, H. H.; Kao, C. H. Purification and Characterization of an Aspartic Protease from the Rhizopus Oryzae Protease Extract, Peptidase R. *Electron. J. Biotechnol.* 2014, *17*, 89.

[44] Kaleas, K. A.; Tripodi, M.; Revelli, S.; Sharma, V.; Pizarro, S. A. Evaluation of a Multimodal Resin for Selective Capture of CHO-Derived Monoclonal Antibodies Directly from Harvested Cell Culture Fluid. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 2014, *969*, 256.

[45] Macauley-Patrick, S.; Fazenda, M. L.; McNeil, B.; Harvey, L. M. Heterologous Protein Production Using the Pichia Pastoris Expression System. *Yeast* 2005, *22*, 249.

[46] Pearson, W. R. An Introduction to Sequence Similarity ("Homology") Searching. *Curr. Protoc. Bioinforma.* 2013, *42*, 3.1.1.

[47] Paul, S.; Lindestam Arlehamn, C. S.; Scriba, T. J.; Dillon, M. B. C.; Oseroff, C.; Hinz, D.; McKinney, D. M.; Carrasco Pro, S.; Sidney, J.; Peters, B.; Sette, A. Development and Validation of a Broad Scheme for Prediction of HLA Class II Restricted T Cell Epitopes. *J. Immunol. Methods* 2015, *422*, 28.

[48] Jawa, V.; Hall, M.; Flynn, G. Evaluating Immunogenicity Risk Due to Host Cell Protein Impurities in Antibody-Based Biotherapeutics. *AAPS J.* 2016, *18*, 1439.

[49] Menegatti, S.; Bobay, B. G.; Ward, K.; Islam, T.; Kish, W. S.; Naik, A.; Carbonell, R. G. C. ". Design

of Protease-Resistant Peptide Ligands for the Purification of Antibodies from Human Plasma. *J. Chromatogr. A* 2016, *1445*, 93.

[50] Gasser, B.; Prielhofer, R.; Marx, H.; Maurer, M.; Nocon, J.; Steiger, M.; Puxbaum, V.; Sauer, M.; Mattanovich, D. Pichia Pastoris : Protein Production Host and Model Organism for Biomedical Research. *Future Microbiol.* 2013, *8*, 191.

[51] Zhang, Y.; Cremer, P. S. Interactions between Macromolecules and Ions: The Hofmeister Series. *Curr. Opin. Chem. Biol.* 2006, *10*, 658.

[52] Omta, A. W.; Kropman, M. F.; Woutersen, S.; Bakker, H. J. Negligible Effect of Ions on the Hydrogen-Bond Structure in Liquid Water. *Science* 2003, *301*, 347.

[53] Batchelor, J. D.; Olteanu, A.; Tripathy, A.; Pielak, G. J. Impact of Protein Denaturants and Stabilizers on Water Structure. *J. Am. Chem. Soc.* 2004, *126*, 1958.

[54] Sripada, S. A.; Chu, W.; Williams, T. I.; Teten, M. A.; Mosley, B. J.; Carbonell, R. G.; Lenhoff, A. M.; Cramer, S. M.; Bill, J.; Yigzaw, Y.; Roush, D. J.; Menegatti, S. Towards Continuous MAb Purification: Clearance of Host Cell Proteins from CHO Cell Culture Harvests via "Flow-through Affinity Chromatography" Using Peptide-Based Adsorbents. *Biotechnol. Bioeng.* 2022, *119*, 1873.

[55] Sánchez-Trasviña, C.; Flores-Gatica, M.; Enriquez-Ochoa, D.; Rito-Palomares, M.; Mayolo-Deloisa, K. Purification of Modified Therapeutic Proteins Available on the Market: An Analysis of Chromatography-Based Strategies . *Frontiers in Bioengineering and Biotechnology* . 2021.

[56] Castro, L. S.; Lobo, G. S.; Pereira, P.; Freire, M. G.; Neves, M. C.; Pedro, A. Q. Interferon-Based Biopharmaceuticals: Overview on the Production, Purification, and Formulation. *Vaccines* 2021, *9*.

[57] Lavoie, R. A.; Chu, W.; Lavoie, J. H.; Hetzler, Z.; Williams, T. I.; Carbonell, R.; Menegatti, S. Removal of Host Cell Proteins from Cell Culture Fluids by Weak Partitioning Chromatography Using Peptide-Based Adsorbents. *Sep. Purif. Technol.* 2021, *257*, 117890.

[58] Lavoie, R.; di Fazio, A.; Blackburn, R.; Goshe, M.; Carbonell, R.; Menegatti, S. Targeted Capture of Chinese Hamster Ovary Host Cell Proteins: Peptide Ligand Discovery. *Int. J. Mol. Sci.* 2019, *20*, 1729.

[59] Ichihara, T.; Ito, T.; Kurisu, Y.; Galipeau, K.; Gillespie, C. Integrated Flow-through Purification for Therapeutic Monoclonal Antibodies Processing. *MAbs* 2018, *10*.

[60] Thomas, G. D. Effect of Dose, Molecular Size, and Binding Affinity on Uptake of Antibodies BT - Drug Targeting: Strategies, Principles, and Applications; Francis, G. E., Delgado, C., Eds.; Humana Press: Totowa, NJ, 2000; pp 115.

[61] Lüdel, F.; Bufe, S.; Bleymüller, W. M.; de Jonge, H.; Iamele, L.; Niemann, H. H.; Hellweg, T. Distinguishing Between Monomeric ScFv and Diabody in Solution Using Light and Small Angle X-Ray Scattering. *Antibodies*. 2019.

[62] Tripathi, N. K.; Shrivastava, A. Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development. *Frontiers in Bioengineering and Biotechnology*. Frontiers Media S.A. December 20, 2019, p 420.

[63] Noh, H.; Vogler, E. A. Volumetric Interpretation of Protein Adsorption: Competition from Mixtures and the Vroman Effect. *Biomaterials* 2007, *28*, 405.

[64] Molden, R.; Hu, M.; Yen E, S.; Saggese, D.; Reilly, J.; Mattila, J.; Qiu, H.; Chen, G.; Bak, H.; Li, N. Host Cell Protein Profiling of Commercial Therapeutic Protein Drugs as a Benchmark for Monoclonal Antibody-Based Therapeutic Protein Development. *MAbs* 2021, *13*.

[65] Bracewell, D. G.; Francis, R.; Smales, C. M. The Future of Host Cell Protein (HCP) Identification

During Process Development and Manufacturing Linked to a Risk-Based Management for Their Control. *Biotechnol. Bioeng* 2015, *112*, 1727.

[66] Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal Sample Preparation Method for Proteome Analysis. *Nat. Methods* 2009, *6*, 359.

[67] Tyanova, S.; Temu, T.; Cox, J. The MaxQuant Computational Platform for Mass Spectrometry-Based Shotgun Proteomics. *Nat. Protoc.* 2016, *11*, 2301.

[68] Marcus, K. *Quantitative Methods in Proteomics*.

[69] Singh, N.; Arunkumar, A.; Peck, M.; Voloshin, A. M.; Moreno, A. M.; Tan, Z.; Hester, J.; Borys, M. C.; Li, Z. J. Development of Adsorptive Hybrid Filters to Enable Two-Step Purification of Biologics. *MAbs* 2017, *9*, 350.

[70] Patil, R.; Walther, J. Continuous Manufacturing of Recombinant Therapeutic Proteins: Upstream and Downstream Technologies BT - New Bioprocessing Strategies: Development and Manufacturing of Recombinant Antibodies and Proteins; Kiss, B., Gottschalk, U., Pohlscheidt, M., Eds.; Springer International Publishing: Cham, 2018; pp 277.

[71] Karbalaei, M.; Rezaee, S. A.; Farsiani, H. Pichia Pastoris: A Highly Successful Expression System for Optimal Synthesis of Heterologous Proteins. *J. Cell. Physiol.* 2020, *235*, 5867.

[72] Heiss, S.; Puxbaum, V.; Gruber, C.; Altmann, F.; Mattanovich, D.; Gasser, B. Multistep Processing of the Secretion Leader of the Extracellular Protein Epx1 in Pichia Pastoris and Implications for Protein Localization. *Microbiology* 2015, *161*, 1356.

[73] Jawa, V.; Terry, F.; Gokemeijer, J.; Mitra-Kaushik, S.; Roberts, B. J.; Tourdot, S.; De Groot, A. S. T-Cell Dependent Immunogenicity of Protein Therapeutics Pre-Clinical Assessment and Mitigation–Updated Consensus and Review 2020. *Front. Immunol.* 2020, *11*, 1.

[74] FDA. Immunogenicity of Protein-based Therapeutics https://www.fda.gov/vaccines-blood-biologics/biologics-research-projects/immunogenicity-protein-based-therapeutics.

[75] Shah, K. A.; Clark, J. J.; Goods, B. A.; Politano, T. J.; Mozdzierz, N. J.; Zimnisky, R. M.; Leeson, R. L.; Love, J. C.; Love, K. R. Automated Pipeline for Rapid Production and Screening of HIV-Specific Monoclonal Antibodies Using Pichia Pastoris. *Biotechnol. Bioeng.* 2015, *112*, 2624.

[76] de Sá Magalhães, S.; Keshavarz-Moore, E. P. Pastoris (Komagataella Phaffii) as a Cost-effective Tool for Vaccine Production for Low- and Middle-income Countries (Lmics). *Bioengineering* 2021, *8*.

[77] Goodrick, J. C.; Xu, M.; Finnegan, R.; Schilling, B. M.; Schiavi, S.; Hoppe, H.; Wan, N. C. High-Level Expression and Stabilization of Recombinant Human Chitinase Produced in a Continuous Constitutive Pichia Pastoris Expression System. *Biotechnol. Bioeng.* 2001, *74*, 492.

[78] Brady, J. R.; Love, J. C. Alternative Hosts as the Missing Link for Equitable Therapeutic Protein Production. *Nat. Biotechnol.* 2021, *39*, 403.

Supplemental Information

Understanding *K. phaffii* (*Pichia pastoris*) Host Cell Proteins: Proteomic Analysis and Flow-through Affinity Clearance

Sobhana A. Sripada[1], Driss Elhanafi[2], Leonard Collins[3], Taufika I. Williams[3], Marina Linova[4], John Woodley[4], Stefano Menegatti[1,2]

1. Department of Chemical and Biomolecular Engineering, North Carolina State University, 911 Partners Way, Raleigh, NC 27695, USA
2. Biomanufacturing Training and Education Center (BTEC), 850 Oval Drive, Raleigh, NC 27606, USA
3. Molecular Education, Technology, and Research Innovation Center (METRIC), North Carolina State University, 2620 Yarbrough Dr., Raleigh, NC 27607, USA
4. Department of Chemical and Biochemical Engineering, Technical University of Denmark, 240, 2800 Kongens Lyngby, Denmark
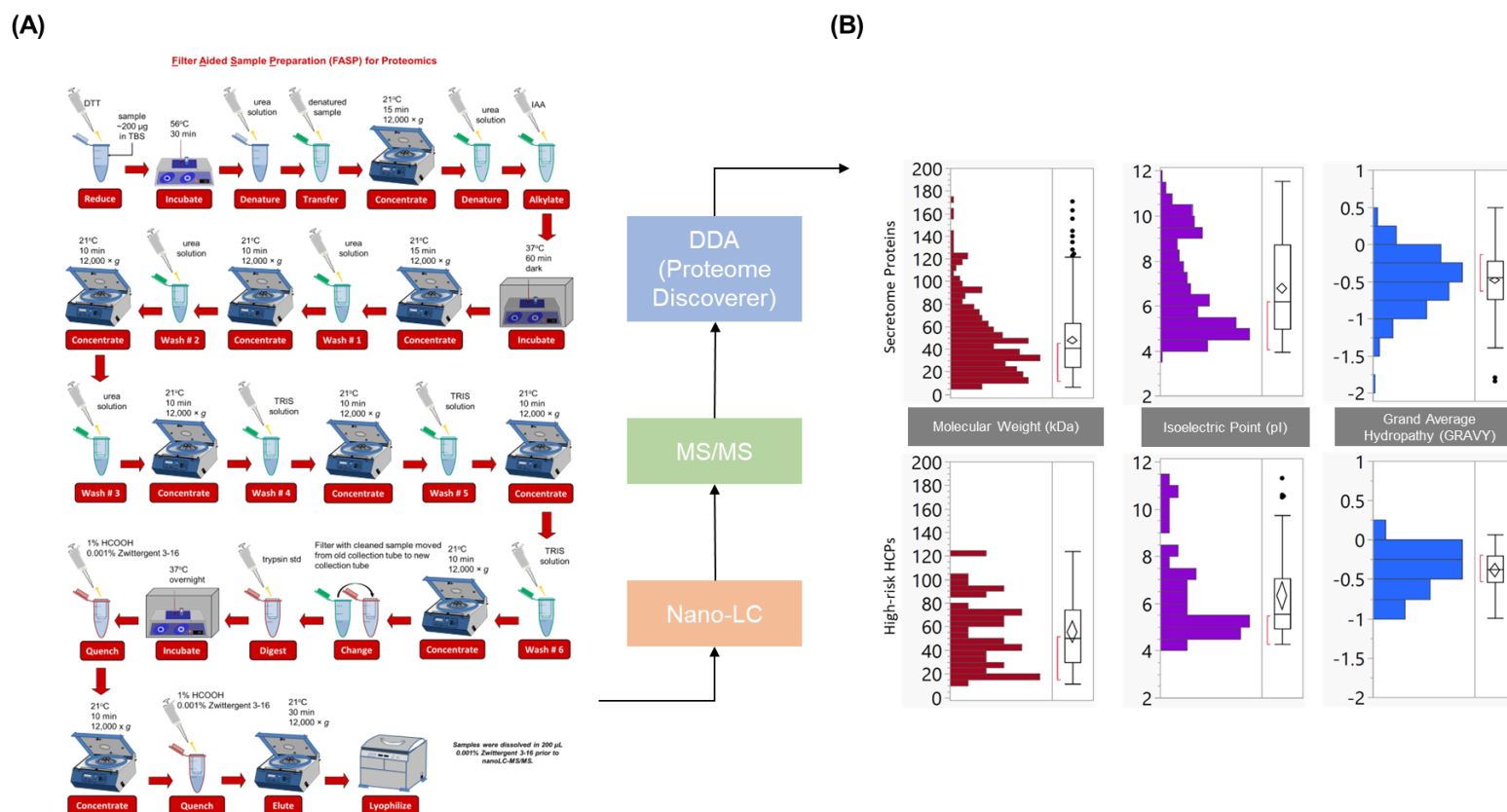
Corresponding author: smenega@ncsu.edu.

(A)

(B)

***Figure S1: (A)** Protocol of Filter-Aided Sample Preparation (FASP) implemented to execute the bottom-up proteomic analysis of K. phaffii CCF and flow-through fractions obtained by loading the CCF on control (CaptoQ, CaptoAdhere, CHO LigaGuard™) and experimental (PichiaGuard-TP50F and PichiaGuard-TP650M) resins; **(B)** Distribution of physicochemical properties (i.e., molecular weight, isoelectric point (pI), and GRAVY indices) of HCPs identified in K. phaffii CCF (top row) and specifically of their "high-risk" subset (bottom row).*
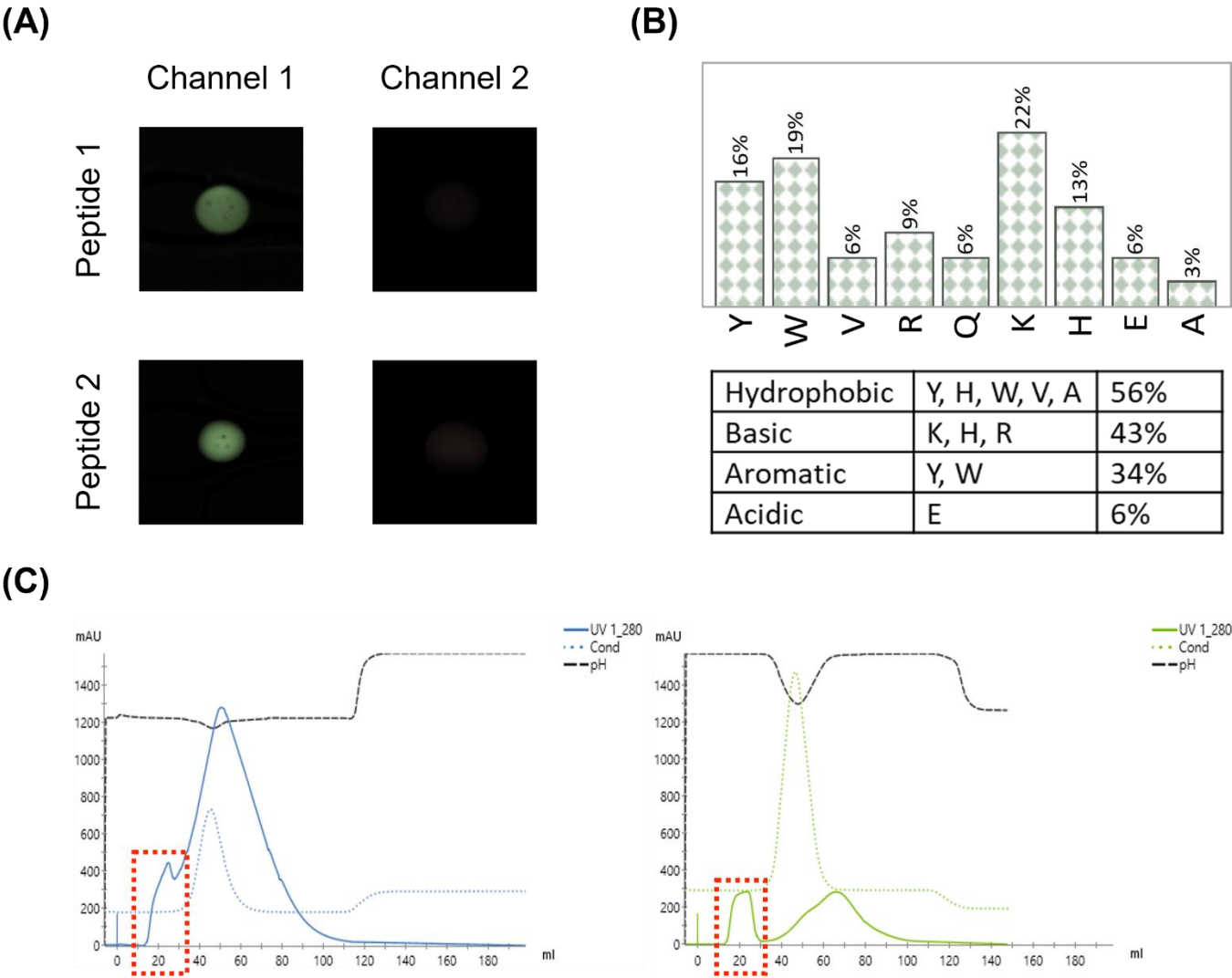
Figure S2: (A) *Visualization of Channel 1 (green, positive for HCP-binding) and Channel 2 (red, positive for polyclonal Ab binding) from the microfluidic peptide library screening setup showing select HMBA-ChemMatrix resin beads coupled with test peptides that had high HCP selectivity; **(B)** Frequency and distribution of amino acids identified in K. phaffii HCP binding peptides and their properties; **(C)** Desalting chromatograms obtained from gel-permeation chromatography conducted to buffer-exchange K. phaffii cell cultures used in the current study, corresponding to clarification of culture with (right) and without (left) protease-inhibitors.*
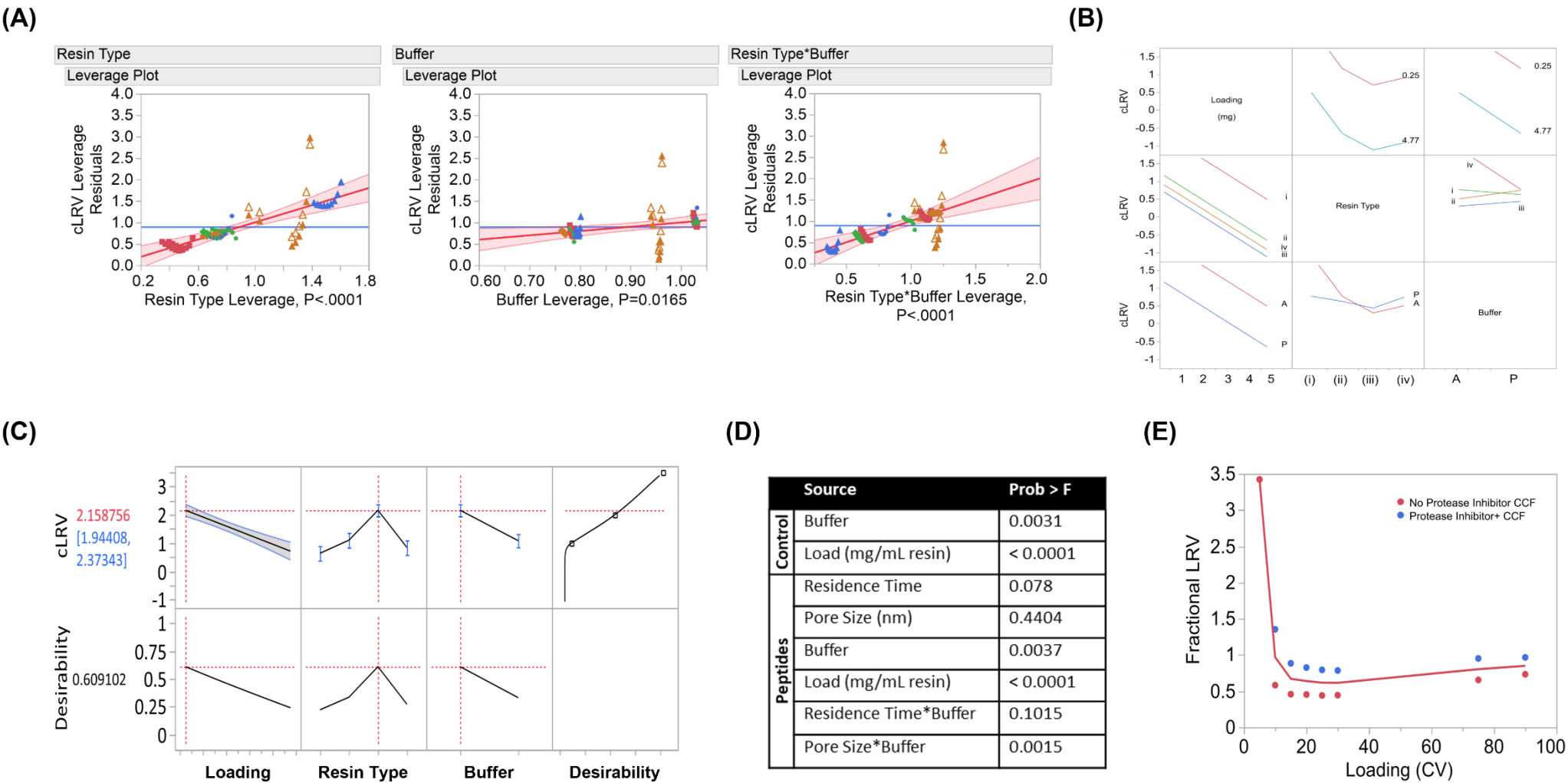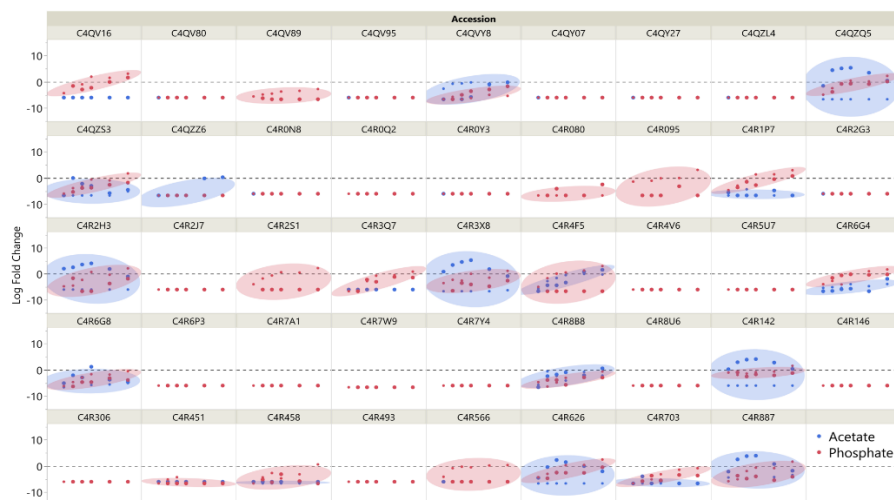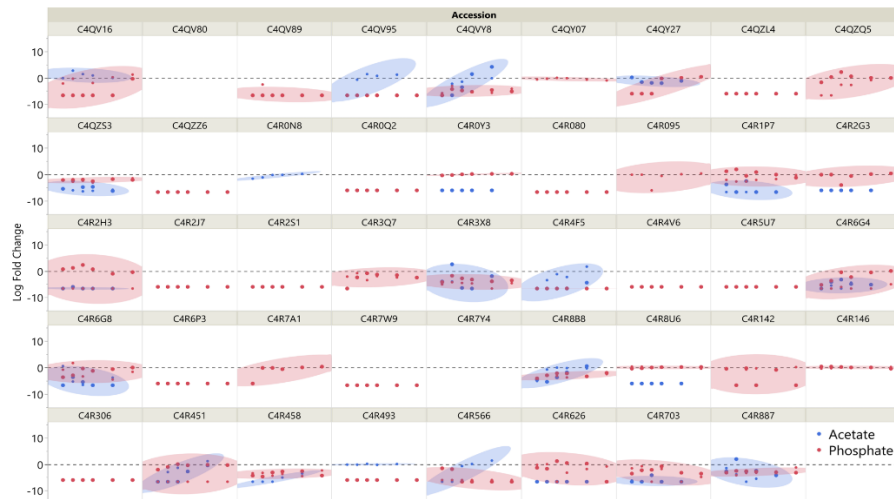
(A)

(B)

(C)

(D)

	Source	Prob > F
Control	Buffer	0.0031
	Load (mg/mL resin)	< 0.0001
	Residence Time	0.078
	Pore Size (nm)	0.4404
Peptides	Buffer	0.0037
	Load (mg/mL resin)	< 0.0001
	Residence Time*Buffer	0.1015
	Pore Size*Buffer	0.0015

(E)

Figure S3: *(**A**) Leverage plots of 'significant' variables generated by the DOE model within the control study group viz., resin, buffer and the interaction term between them; (**B**) Interaction plot showing trends between independent variables and HCP cLRV observed across performed experiments where A,P represent acetate and phosphate buffers, (i)-(ii)-(iii)-(iv) represent PichiaGuard, CHO LigaGuard, CaptoAdhere and CaptoQ respectively; (**C**) Prediction profiler showing the desirability of choosing PichiaGuard resin operated with Acetate buffer for achieving a goal of maximizing cLRV; (**D**) F-statistics of variables tested in the DOE mode that show buffer type and loading influence HCP clearance significantly; (**E**) fractional log reduction of HCP observed with PichiaGuard CCF treated with and without protease inhibitors.*
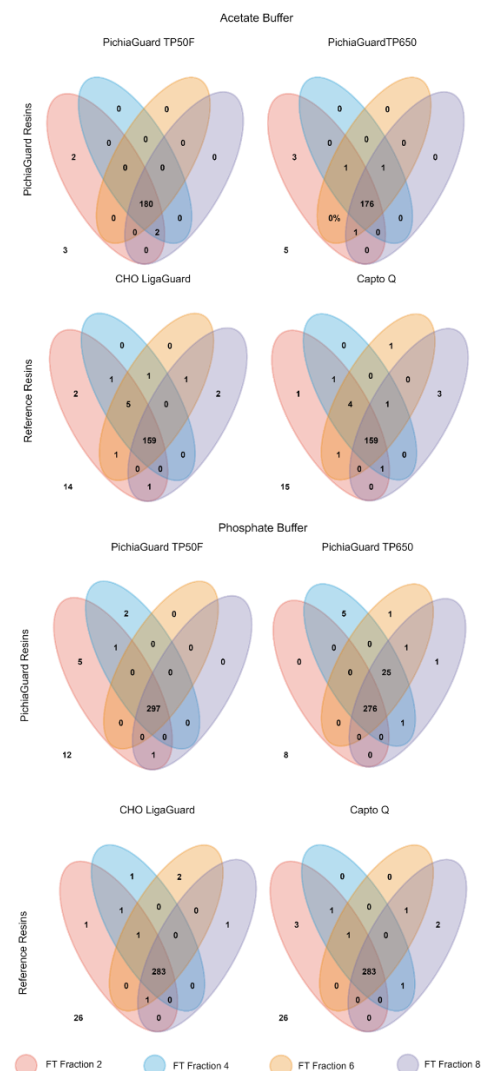
Figure S4: (A, B) *trends of individual high-risk HCPs across different flow-through fractions tested in both reference and PichiaGuard groups respectively, across both acetate and phosphate buffer groups. Within each square, the smaller marker represents groups CHO LigaGuard (top) and PichiaGuard TP-50F (bottom) and the larger markers CaptoQ (top) and PichiaGuard TP-650M (bottom); **(C)** Total number of bound HCPs identified in each flow-through fraction via nano-LC-MS/MS within each experimental condition performed across resin type and buffer condition.*
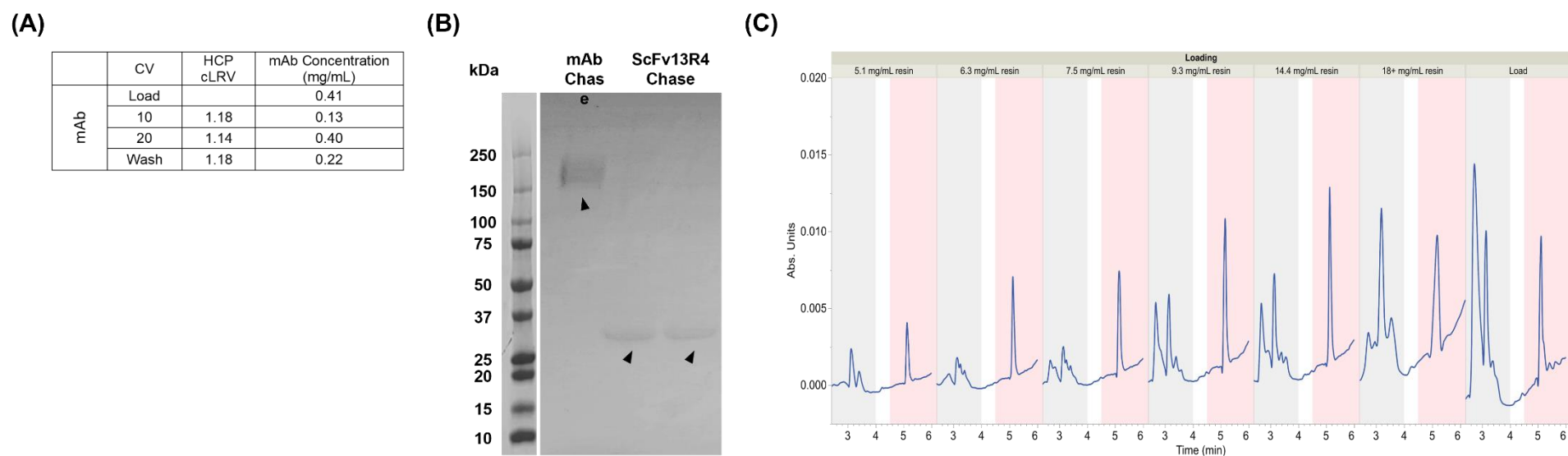
(A)

	CV	HCP cLRV	mAb Concentration (mg/mL)
mAb	Load		0.41
	10	1.18	0.13
	20	1.14	0.40
	Wash	1.18	0.22

(B)

(C)

Figure S5: (A) *Yield and purity observed in mAb purification from K. phaffii CFF; **(B)** SDS-PAGE of high-salt chase fractions post PichiaGuard processing showing a small degree of product binding to the column, possibly due to association with other HCPs at run conditions; **(C)** Reverse phase chromatograms of individual flow-through fractions obtained from ScFv13R4 purification study compared at different intervals of protein loading (mg protein/mL resin) where we observe changing ratios of impurity peaks (grey) and the product peak (peach). The ratios reported in this article have been calculated by dividing areas under the curves between these retention times respectively.*