Development of Yeast-displayed scFv and Monoclonal Antibodies for *Entamoeba histolytica* Cyst Detection

Lauren Jane Spadafora

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Gerard Cangelosi, Chair

Marilyn Parsons

Paul Lampe

Program Authorized to Offer Degree:

Pathobiology

University of Washington

Abstract

Development of Yeast-displayed scFv and Monoclonal Antibodies for *Entamoeba histolytica* Cyst Detection

Lauren J. Spadafora

Chair of the Supervisory Committee:
Gerard A. Cangelosi, Ph.D.
Pathobiology Program, Department of Global Health

The aims of this dissertation research are two-fold: 1) develop an accelerated affinity reagent pipeline that overcomes the limitations seen in utilizing recombinant antibodies and *in vitro* display systems; and 2) validate a novel *Entamoeba histolytica* cyst biomarker in stool, utilizing reagents obtained in the accelerated pipeline as well as by classical means. As the human proteome remains in focus and as new and remerging disease threats continue to exact terrible burdens worldwide, there is an enormous need for antigen-specific affinity reagents that can be produced rapidly and with little cost. The most widely used affinity reagent, the monoclonal antibody, is generated through a months-long, laborious, and expensive methodology, with no guarantee of a successful output. In response, a variety `of antibody-derived molecules have been generated recombinantly and have been displayed on a wide number of cells and organisms as combinatorial libraries for rapid, easy isolations of desired affinity reagents. However, removing these reagents from their display environments, the environment in which they were selected to function, often results in inactivated and non-functional product. Here, we overcome this limitation by, first, stabilizing selected yeast-displayed recombinant antibody (called yeast-scFv) via lyophilization and by, secondly, creating

a novel class of affinity reagent, called nanoyeast-scFv. Nanoyeast-scFv are simply generated by crudely fragmenting whole yeast-scFv into cell wall fragments that still display functional antibody. These reagents demonstrated superior sensitivity (in pg/mL to ng/mL range) and specificity to their antigens when tested in various solid support assay formats.

Entamoeba histolytica is a protozoan pathogen that is difficult to detect, due to the lack of a sensitive and affordable diagnostic test and to the presence of morphologically identical commensal amoeba in endemic areas. Current stool antigen detection tests target the trophozoite form, which is typically not plentiful or stable in stool. Our strategy was to validate a novel biomarker of the organism's cyst stage. About 50 individual *E. histolytica* proteins were rationally selected for further investigation by filtering through recent cyst proteomic and transcriptomic data and identifying *E. histolytica*-specific sequences. As *E. histolytica* cysts cannot be obtained *in vitro*, these targets were recombinantly expressed and subjected to yeast-scFv selection, as well as traditional monoclonal antibody production. A monoclonal antibody to a known cyst wall protein, the Jacob2 lectin, was found to bind fixed, *E. histolytica* cysts from xenic culture specifically. This is one of very few times that a monoclonal antibody to *E. histolytica* cysts has been identified, as well as the first time a monoclonal antibody specific to such cysts in a fixative has been reported.

Supplemental Files (Excel):

Table 2S1: Abundant *E. histolytica* Biomarkers (LC-MS/MS-identified in ≥3 of 5 Stool Samples)

Table 2S2: LC-MS/MS-identified *E. histolytica* Biomarkers With Upregulated or Unchanged mRNA Transcript in Encysting Cultures Versus Axenic Cultures

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

It is only appropriate that I begin with acknowledgements, as completing this work has been far from an isolated journey. I am truly blessed and graced in this life, and it has been an incredible opportunity and privilege to continue my education.

I am extremely grateful to Dr. Jerry Cangelosi for his mentorship over these five years. I was eager to work on a translational research project when arriving in Seattle, and I will always be thankful that he gave me a chance to do so. His patience, kindness, and enthusiasm made working for his group an excellent learning experience and a joy.

I don't know what I would do without a supportive and fun-loving group of labmates. I am forever grateful to Kris Weigel for the many hours in which he lent an ear and gave advice, whether it was about experimental design or day-to-day life. I want to thank Annie Becker for coordinating the sample collections in this project, for helping me with administrative work, for training me, and for bringing an unparalleled energy to the lab. Dr. Sean Gray was instrumental in getting me started on this project, passing down his expertise in protein expression and yeast display, and I also enjoyed briefly working with Dr. Brian Oliver, a great source of advice on cloning and on adapting to Seattle as an Ohioan. Dr. Sara Murray not only helped with this project for a year, but kindly proofread application essays and encouraged me during a time in which I was struggling. As a program manager, Dr. Maegan Ashworth Dirac got the AMPP project on a good start before I joined the lab, and I am appreciative of her advice and her feedback on presentations. I've also had a stellar group of lab assistants and rotation students: Joji, Jingwen, Moira, Justine, and Jay, thank you for being gracious and fun mentees! Also many thanks to Kelly Jones, Hung Nguyen, Herakles Li, Rachel Wood, Felicia Nguyen, Connie Tzou, and Jenny Lohmiller for the camaraderie and the laughter.

To all of the AMPP partners at ICDDR,B, University of Queensland, University of Virginia, and PATH: thank you for your hard work and support over these years. I could not have asked for better collaborators! I am also indebted to the SSGCID and for the FHCRC Antibody Technology Core for productive and positive collaborations and to Seattle Biomed (now CIDR) and UW DEOHS for making the lab a congenial and upbeat place to work.

I am especially grateful for all human subjects who donated samples for this project.

Additionally, I would like to thank my doctoral committee- Dr. Marilyn Parsons, Dr. Paul Lampe, Dr. Peter Myler, and Dr. Barry Lutz- for their mentorship, which helped me to improve my critical thinking and communication skills. Because of them, I leave now a better scientist. I extend additional thanks to Jerry, Dr. Parsons, and Dr. Lampe for reading this dissertation.

I am also indebted to the Pathobiology program- first to Dr. Lee Ann Campbell, Rachel Reichert, and Ashley Ziegler for their support throughout the years. Many thanks to my fellow Pathobiology students for their friendship and insight throughout this journey!

To Drs. Michael Crowder, Mitch Balish, Martha Castañeda, Sheila Lukehart, Lorenzo Giacani and Arturo Centurion-Lara, thank you for your mentorship and support. I could not have done this without you!

I am incredibly thankful for my family and friends, especially my parents Frank and Suzanne (who hauled my Chevy Cavalier cross-country in three days flat), my BMB friends for their love and epic adventures, my church friends who never ceased to lovingly remind me of Phil 4:5-6, and my roommates Alison, Ginny, and Kim for walking with me during the highs and the lows.

Finally, to Chris Gable- thank you for waiting for me for five years. God has truly has blessed me with you. I can’t wait to see what new adventure is waiting for us.

LIST OF FIGURES

LIST OF TABLES

CHAPTER I: INTRODUCTION

I. THE RATIONALE FOR AN ACCELERATED PIPELINE FOR AFFINITY REAGENTS

The biological world is driven by countless interactions of biomolecules, and exploiting these interactions for biomedical research and development has garnered increased interest, particularly in elucidating the human proteome[1–4]. Central to this endeavor are affinity reagents, defined as molecules that recognize and bind to a target of interest[3]. Of the several classes of existing affinity reagents, antibodies are the most widely used today[3–5].

Traditional Antibody Affinity Reagents and Their Production

Antibodies have been harnessed for their natural ability to recognize non-self material *in vivo*. These proteins are first produced and displayed on the surface of B cells, and, when they bind very specifically to a foreign material (an antigen), the B cell is signaled to complete a number of immune functions. One such function is to differentiate into a plasma cell, which secretes the soluble antibody to aid in neutralizing and clearing the recognized non-self antigen[6–8]. **Figure 1-1** is a representation of antibody structure found in humans, as well as laboratory animals such as mice, rabbits, and goats. An antibody is comprised of two sets of paired heavy and light polypeptide chains, which interact via noncovalent interactions and inter-chain disulfide bonds[9,10]. The chains themselves are comprised of immunoglobin domains that each consist of beta sheets and an intra-domain disulfide bond[9,10]. In placental mammals, antibodies can be classified into five classes by the sequences of their heavy chain C-termini, and these

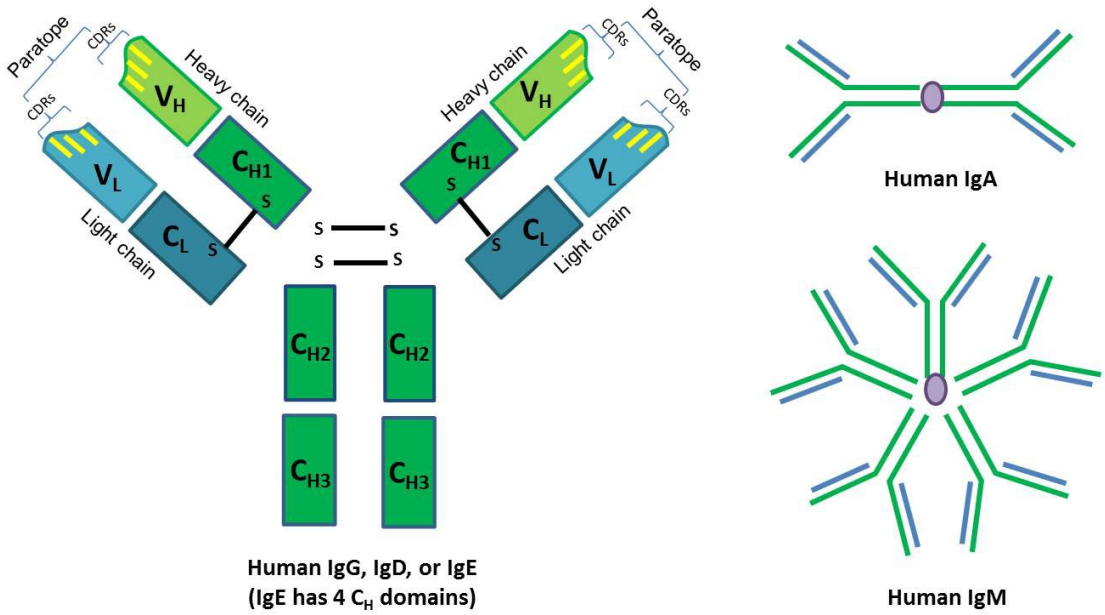

Figure 1-1. The structure of human antibodies. A single molecule is comprised of four polypeptides, two heavy chain copies two light chain copies, which are bound together by disulfide bonds. The polypeptides' variable and constant domains (squares, V=variable, C=constant) are comprised of beta sheets and an intra-domain disulfide bond ("S—S"). A heavy and light chain pair to form an antigen binding pocket, called a paratope. Diverse paratopes are conferred by hypervariable complementary determining regions (CDRs, yellow stripes) that are located within the N-terminal variable domains (labeled V_H and V_L). An antibody's class is determined by the heavy chains' constant domains (labeled C_{H1-3}). The IgA, IgD, and IgE isotypes are oligomeric. Linked by a J chain (purple oval), isotype IgA is typically dimeric, and IgM is pentameric.

classes differ in tissue distribution, immune function, and propensity to multimerize[11]. Meanwhile, the source of the antibody repertoire's diversity and wide recognition lies within the N-terminal domains of these chains, where six regions of sequence variability, called complementary determining regions (CDRs), form a unique antigen binding pocket, the paratope[10,12–14].

There are a number of significant mechanisms that contribute to the diverse antibody repertoire. A first mechanism is antibody gene arrangement and rearrangement. The variable domain-encoding genes are divided into two types of coding segments (V and J) on light chain and three types (V, D, J) on heavy chain[11,15,16]. In humans, there are approximately 50 V_H, 20-30 D, and 6 J_H functional heavy chain segments, as well as 35-40 V and 4-5 J segments for each light chain, encoded in germline DNA[17–24]. During B cell development, single segments of each

type are recombined together, along with the C segment encoding the antibody constant region, to form a complete gene[15,25–27]. A second mechanism providing diversity is the recombination of the antibody gene segments, which occurs imprecisely and can involve nucleotide insertions and deletions[15,28]. While CDR1 and CDR2 of each chain are encoded in the germline V segments, CDR3 regions lie in the V(D)J junctions and have the greatest sequence variability[14,16]. Finally, diversity is increased during affinity maturation, in which somatic hypermutation occurs within CDRs to create and select for B cells with heightened antigen affinity[29,30].

In the late nineteenth century, antibodies were first observed at work in sera of animals immunized with particular pathogens or toxins, aptly named "anti-sera" or "anti-toxins"[31,32]. Disease protection conferred by these fluids was suspected to happen via very specific, "lock-and-key" style interactions of the toxins and "side-chains", molecules present on cell surfaces but later secreted into the bloodstream[33,34]. The first antibody-based diagnostic tests, in which the combination of antigen and homologous immune sera produce a visible reaction, were reported by Gruber in 1896 (agglutination) and Kraus in 1897 (precipitation)[35–37]. In subsequent decades, antibodies were purified and identified as protein molecules, the specificities of antibodies were further revealed by immunizations with hapten-modified antigens, and, ultimately proving the "lock-and-key" idea, Linus Pauling showed that antigen-antibody interactions were due to structural complementarianism[38–41]. Several more technologies exploiting antigen-antibody interactions have since been developed and have gained popularity in research and medicine, including western blots, enzyme-linked immunosorbent assay (ELISA), immunofluorescence microscopy, and flow cytometry[1,42–46].

Antiserum contains polyclonal antibody, a heterogeneous antibody mixture that originates from multiple B cells and that can typically recognize multiple epitopes within a single antigen[6].

While it can often detect antigen more sensitively than a single species of antibody, polyclonal antibody is naturally more prone to cross-react with homologous proteins, which hinders accurate target detection[6,47,48]. Furthermore, a working polyclonal product is of limited clinical laboratory utility, because the supply varies in activity and composition from animal to animal, lot to lot[6,48,49]. In the 1970s, Köhler and Milstein revolutionized the affinity reagent landscape by developing production of monoclonal antibodies[50]. In contrast to polyclonal reagents, monoclonal material contains copies of a single antibody molecule and hence will bind specifically to a single epitope[6,49,50]. To produce monoclonal antibodies, an antigen is injected into an animal to generate antigen-binding B cells in the spleen or lymph nodes, and the resulting immune cells are fused with immortal myeloma cells to create a hybrid cell line that indefinitely secretes antibody[11,50,51]. Once a single "hybridoma" clone is isolated, an ongoing supply of the monoclonal antibody is assured [6,49].

Decades after this discovery, monoclonal antibodies are widely used, but they have limitations that have frustrated biomedical research efforts. The production of monoclonal antibodies is considered more technically challenging than polyclonal antibody production, labor-intensive, and not high throughput, with lead times for a reagent being up to 1 year[5,6]. In addition, current research and development endeavors are now investigating tens to thousands of biological markers at a time, and developing a monoclonal antibody to each target is financially and logistically unfeasible for most laboratories[2,5]. Furthermore, with the investments of time and money required, monoclonal antibody production can be very risky: resulting outputs can be biased towards immunodominant epitopes, have weak binding affinities, and or have poor, nonspecific, or unreproducible activity in desired applications and tests[5,49,52–54]. Finally, DNA

sequences of monoclonal antibodies are not usually obtained or provided in the process, so it is difficult to genetically engineer isolated antibodies for improved characteristics.[4,5]

Alternative Antibodies and Display Systems

In response to the limitations of monoclonal antibodies, investigations have turned to recombinant antibody selection and expression in microorganisms and cell lines. A key accomplishment toward this endeavor was PCR-amplification and cloning of antibody-encoding genes[55–57]. This allowed for *in vitro* synthesis, mutagenesis, and recombination of variable domains to confer desired specificities and affinities[58–60]. It also led to the production of antibody analogues that retained parental characteristics but were smaller and easier to manipulate (**Figure 1-2**). The first demonstrations of recombinant antibody expression were production of the Fab antibody fragment in *Escherichia coli,* followed by a study detailing secretion of Fabs in a

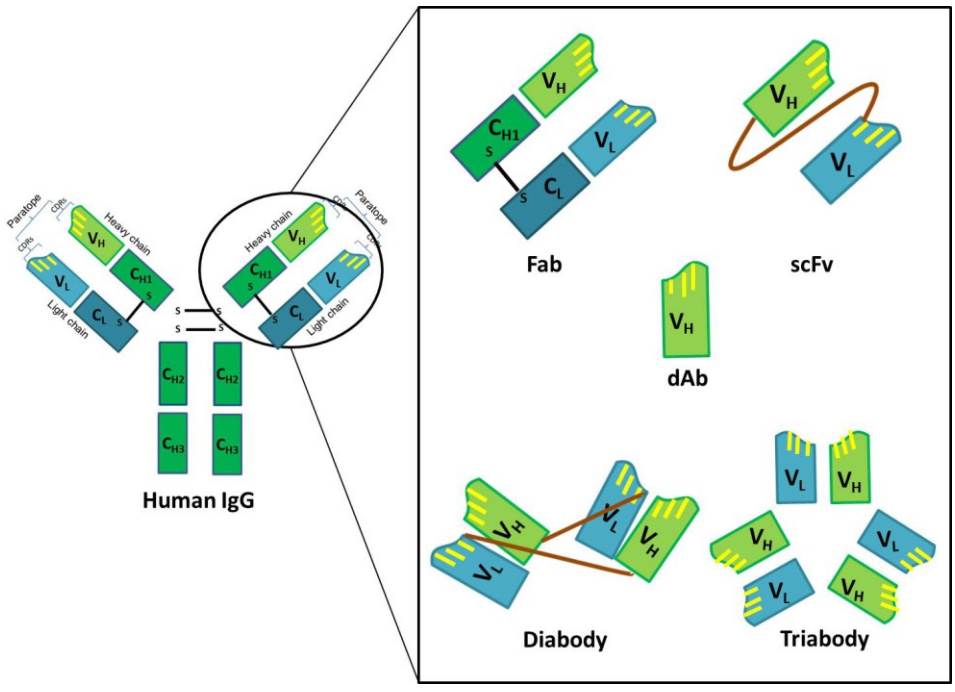

Figure 1-2. Illustrations of recombinant antibody fragments, which are derived from antibodies' N-terminal domains (circled). CDRs and disulfide bonds are marked as before (yellow stripes and "S—S"). Genetically engineered flexible linkers that connect recombinant antibody domains in scFv and diabodies are represented by dark orange lines.

Saccharomyces cerevisiae system[61,62]. Soon after, the fragment variable (Fv), comprised of naturally associating heavy and light variable domains, and the single chain Fv (scFv), the same domains fused together by a flexible linker, were similarly expressed in recombinant microorganisms[63–66]. Another report soon following demonstrated the production of single but still active V_H and V_L domains (named dAbs) in *E. coli*[67]. Lastly, shortening of scFv linkers resulted in scFv pairing into diabodies, and further shortening or complete elimination of the linker created stable triabodies[68,69].

The experiments described above facilitated more rapid engineering and production of antibodies, but they were limited in that they began with a traditional antibody of known specificity. Methods to create and screen vast repertoires of this new antibody class were needed. Not long after the recombinant expression systems were reported, large combinatorial Fab libraries of murine V_H and V_L domains (10^7 diversity) were generated in phage lambda and could be screened for antigen-specific binders via plaque assays[56,70,71]. A single scFv, shortly followed by a combinatorial library, was also successfully displayed on the surface of M13 phage by fusion with gene III protein, and further experiments showed that rare, antigen-binding phage could be enriched out of millions irrelevant phage by antigen affinity capture[72–74]. Since then, Fabs and diabodies have also been expressed and selected via phage display, and libraries have been constructed from the B-cell mRNA of target-immune and non-immune humans and animals[75–79]. Further development of phage display technology has also produced larger libraries (up to 10^{11} antibodies), various selection strategies, and *in vitro* affinity maturation methods to increase recombinant antibody affinity[14,79]. A large benefit of phage display and resultant libraries is that, once a large library is constructed, antigen-specific reagents with subnanomolar affinities can be selected in little as two weeks[80].

In 2003, heterologous peptide display on the agglutinin surface receptor of *Saccharomyces cerevisiae* was harnessed to build a non-immune scFv library of 10^9 diversity[81,82]. Yeast-displayed antibodies libraries have typically been less diverse than phage-displayed libraries due to limited transformation efficiency, but yeast display has several advantages over phage display[14,83]. First, in contrast to the requirement for *E. coli* to propagate antibody-displaying phage, yeast possess machinery capable of correctly folding and modifying mammalian-derived proteins and do not introduce expression bias during selections[81]. Indeed, investigators conducting parallel phage and yeast display selections of the same antibody library found that the yeast system produced a wider diversity of specific antibodies[84]. Another clear advantage of yeast over phage display is the ability to couple it with fluorescently activated cell sorting (FACS) methodology, which enables more discriminate selection and enrichment of antigen binding cells[82,85,86]. Yeast can bind monomeric antigen in suspension, which helps to overcome avidity effects observed in phage panned against immobilized antigen, and, like phage display, yeast display selections can be completed in as little as two to three weeks[82] . Finally, out of the currently used display systems, it was yeast display, coupled with *in vitro* affinity maturation, that has produced the strongest antibody affinity reagent to date, with Kd as low as 48 fM[87]. Other systems utilized for display of recombinant antibody include bacterial display (*Escherichia coli, Staphylococcus aureus, Bacillus* spores), viral display, mammalian display, and a cell-free method that combines expressed antibody with RNA transcript via ribosomes complexes[88–96].

Despite all of these elegant breakthroughs, alternative antibodies have yet to become a viable alternative to animal-derived antibodies. Currently, there is a bottleneck in the subcloning and expression of displayed reagents as soluble molecules. From studies of recombinant

antibody expression in *E. coli*, which has been popular due to its ease and affordability, many factors have been implicated in insoluble and inactive expression, including antibody and scFv linker primary sequence, proteolytic digestion of product, aggregation via exposed, hydrophobic areas within the domains, and destabilization and misfolding of product when expressed in reduced, cytoplasmic environments[97]. In yeast, components of protein folding and trafficking through the ER and Golgi can influence an antibody's ultimate yield[98–100]. Additionally, it has been noted that higher thermodynamic stability correlates with higher protein display on yeast and with higher secretion yield[101]. Overall, as for any recombinantly expressed protein, the particular host systems and expression conditions required to maintain activity are unique for every antibody and must be screened and optimized[97,102]. The central and critical issue underlying these observations is summarized as thus: removing selected antibody from display removes it from the environment in which it was selected to function. For yeast display, such environment is the yeast cell wall, where the antibody's agglutinin receptor fusion partner and other cell wall components may influence the antibody's folding and stability.

Without considering these pitfalls when selecting or constructing a display library, optimizing antibody expression can become another long and laborious process, ultimately not providing much benefit over traditional antibody generation. As described later in this chapter, an aim of this dissertation is to overcome these limitations of antibody display.

II. THE RATIONALE FOR AFFINITY REAGENT TARGET: ENTAMOEBA HISTOLYTICA

Epidemiology of *E. histolytica*

Entamoeba histolytica is a protozoan human parasite and the causative agent of amoebiasis. It was first reported by Losch in 1875 and named by Schaudinn in 1903 for its ability to lyse cells[103]. In the 1980s and 1990s, it was estimated that 30-50 million infections and 100,000 deaths worldwide per year were attributable to this organism[104,105]. Between 1990 and 2007, 134 deaths in the United States were attributable to amoebiasis[106]. Clinical manifestations of amoebiasis arise in 10-25% of infections and can include diarrhea, dysentery, amoeboma, and toxic megacolon[107–109]. A very small proportion of infected individuals may later develop extraintestinal abscesses in the liver, brain, and lung, often without bowel symptoms or positive stool detection[110,111]. Asymptomatic infections in adults can last at least 12-15 months before clearance, and, in children in an endemic setting, infections typically lasted two to three months[108,109,112].

E. histolytica is transmitted through the fecal-oral route, primarily via contaminated food and water with some reports of sexual transmission[111,113,114]. Infection and disease are mostly found in areas of poor sanitation[114]. Cases in the developed world are often in individuals recently traveling from endemic areas and their close contacts, and outbreaks have occurred during failure of water infrastructure and in institutional settings[115–118]. Increased prevalence of *Entamoeba* infections have also been noted amongst men who have sex with men[119]. Populations at risk of developing severe disease include adults ≥40 years old, pregnant women, immunocompromised individuals, and patients taking corticosteroids[111,114,120]. The negative impact on children is also

clear: diarrhea is a leading cause of death under age five globally, and enteric disease in general is associated with stunted growth, malnutrition, and impaired cognitive function[121,122].

Life Cycle and Biology of *E. histolytica*

Humans are the only known host of *E. histolytica*[114]. The organism has a dual-stage life cycle, in which it will enter and eventually exit the body as a round, impermeable cyst with a 10-15 micron diameter[111,114]. When cysts enter the gastrointestinal tract, they survive passage through the stomach and begin the excystation process in the intestine[111,114,123]. A lectin with high affinity to galactose and N-acetyl-galactosamine (henceforth called Gal/GalNAc lectin) allows trophozoites to adhere to the glycosylated mucin layer of intestinal epithelial cells in the colon[123].

Pathogen factors predictive of a symptomatic infection have yet to be completely elucidated. Disease occurs when colonizing *E. histolytica* trophozoites stimulate a strong pro-inflammatory response, marked by neutrophil influx, and invade the intestinal epithelium[124]. Once in contact, invasive strains are able to kill a number of human cell lines and immune cells, including PMNs, T-cells, and macrophages, via calcium influx, dephosphorylation, and apoptosis in the host cell[123,125]. As demonstrated by competition studies, the Gal/GalNAc lectin is necessary for contact and subsequent cell killing, but it is not the effector molecule that causes the death[123,125]. Recently, it has been shown that trophozoites can kill epithelial cells by ingesting pieces of the cell until it dies, a process known as trogocytosis[126]. In contrast to nonpathogenic species *E. dispar*, *E. histolytica* expresses higher levels of cysteine proteases, which degrade mucin and antibodies and stimulate pro-inflammatory signals, and of peroxiredoxin, which is one of the parasite's few antioxidant enzymes that can combat host reactive species[124,127,128].

Additionally, the organism possesses pore-forming proteins ("amoebapores") that are necessary for liver abscess development, though their mechanism of action is still unclear[125,129].

Host factors important for resistance against *E. histolytica* are still under investigation. Though the acute, neutrophil-rich response exacerbates intestinal damage, a pro-inflammatory Th1 response is required to effectively clear persistent amoeba, whereas an anti-inflammatory Th2 response is non-protective, likely due to suppression of IFN_y[124]. Development of such adaptive responses may be influenced by an individual's MHC alleles and their presentation of certain peptides during T cell activiation[124]. Production of anti-trophozoite serum IgG was associated with increased risk of infection and diarrhea and was observed in families, suggesting that susceptibility to the parasite can be influenced by host genetics[112,130]. Interestingly, this non-protective response often contained IgG4 isotype antibody, which would have been produced during a Th2 response[112,124]. Meanwhile, mucosal IgA antibodies developed against the Gal/GalNAc lectin have been identified with protective properties, but they are not long-lasting[109,112,130]. Additionally, host IL-10 production may play a role in resisting amoeba invasion, by maintaining intestinal barrier integrity and repressing pro-inflammatory responses[124]. Though invasive strains are equipped with several potential mechanisms of evading it, complement has also shown to be protective, and, interestingly, complement in female human sera was more effective at lysing trophozoites than that in males, who have a higher risk of liver abscess[123,124,131]. Very recently, a cohort study demonstrated an association of leptin receptor polymorphisms and risk of infection; receptor signaling in intestinal epithelial cells was soon after shown to be important, likely leading to epithelium regeneration and resistance to apotosis[132,133]. Finally, composition of the gut microbiome may be of importance, as

certain bacteria have been implicated in upregulating adherence, downregulating adherence, or inducing encystation[134–136].

Researchers have long been able to excyst *Entamoeba* organisms and propagate trophozoites in culture, but a reliable and efficient method to encyst trophozoites *in vitro* is yet to be established[137,138]. Xenic cultures, in which *Entamoeba* isolates are co-cultured with fecal bacteria, have produced low levels of spontaneously encysting trophozoites[137,138]. Meanwhile, attempts to induce stage conversion on a larger scale are numerous, but there is worry that the cysts created via these protocols are not entirely authentic[136,139–141]. Utilizing reptilian parasite *Entamoeba invadens* as an in vitro model, encystation has occurred following trophozoite aggregation via galactose ligands, osmotic shock, or glucose depletion[136,142]. Additionally, investigation of *E. histolytica* xenic cultures suggests interaction with bacterial flora is important for this process[136]. *E. histolytica* cysts are believed to have a cell wall structure in line with a "wattle and daub" model (**Figure 1-3**)[143,144]. Upon trigger, a trophozoite will begin secretion of cell wall materials via vesicles. Chitin is deposited to the plasma membrane via large vesicles, with cross-linking Jacob glycoproteins arriving via separate, smaller vesicles and anchoring via

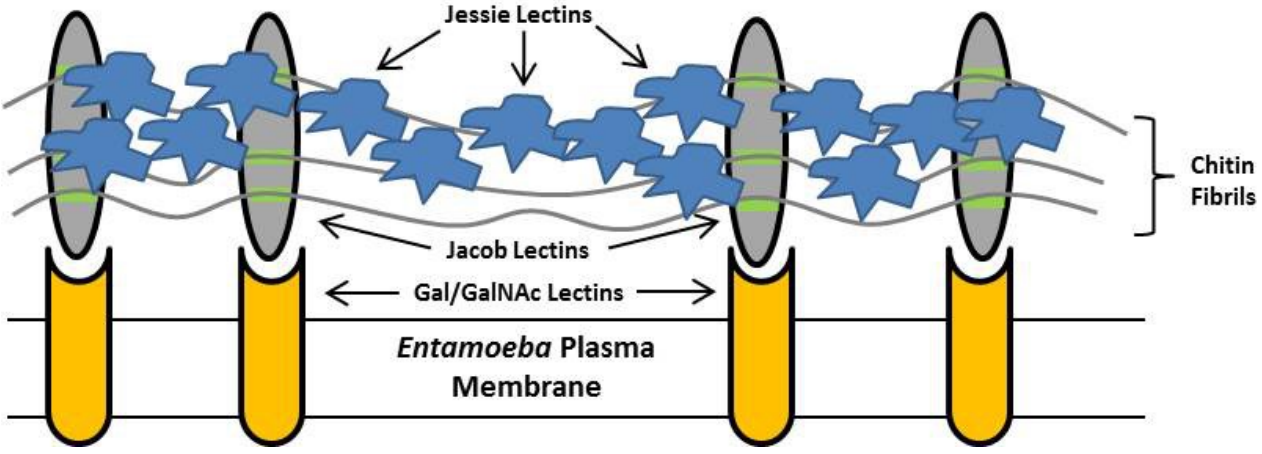

Figure 1-3. The structure of the *Entamoeba* cyst cell wall, as described by the "wattle and daub" model[143,144]. Chitin and glycosylated Jacob lectins are deposited in separate vesicles during early and middle stages of encystation. The Jacob lectins bind back to Gal/GalNAc lectin and crosslink synthesized chitin fibrils via chitin-binding domains (green stripes). In the last stage of encystation, Jessie lectins cover the forming wall and provide a final seal. This figure is adapted from [143].

the Gal/GalNAc lectin[143–146]. In the final stages, Jessie lectin completes the wall-off of the cyst as the daub, and henceforth, nothing other than water and small molecules enter[143].

Treatment and Management of *E. histolytica* Infection

E. histolytica infection and disease is curable, and recovery from liver abscess, once largely a fatal manifestation, is achievable with timely detection and treatment[114]. The tissue amoebacide metronidazole is prescribed for invasive disease and liver abscess, but it is not effective in resolving asymptomatic infections, as it is depleted during its travel to the bowel and by commensal gut bacteria[104,110]. The luminal amoebacide paramomycin is useful for asymptomatic infections[104,147]. Unfortunately, all *Entamoeba* drugs have gastrointestinal side effects, but new medicines are being pursued for more tolerable regimens and for prevention of drug resistance[111,148].

A vaccine against *E. histolytica* has yet to be developed, but it has been a topic of interest for decades. Immunization of animals with Gal/GalNAc lectin has shown protection, but human trials have yet to be conducted[149].

Diagnosis of *Entamoeba histolytica* Infection

Despite adequate treatments, *E. histolytica* remains a problem throughout the world due in part to limitations in its detection. As far as clinical diagnosis, colitis occurs with insidious and non-specific symptoms, whereas dysentery can also be caused by inflammatory bowel disorders and several bacterial species[111]. Amoeba can be collected in stool and viewed by microscopy, but the procedure has poor sensitivity and specificity relative to other methods[150–153]. Some issues that can complicate stool collection and result in missed cases are the intermittent shedding of

organisms and the rapid degradation of motile trophozoites in stool[105,150]. Another severely limiting problem of microscopy was the confirmation that *E. histolytica* strains previously thought to be "nonpathogenic" were indeed two additional but morphotypically identical *Entamoeba* species, *E. dispar* and *E. moshkovskii*[154,155]. Distinguishing pathogenic *E. histolytica* cells from non-pathogenic *Entamoeba* species is a critical challenge.

Serological detection of *Entamoeba* was once praised as a helpful method of detection. It has been long known that individuals infected with *E. histolytica*, regardless of clinical manifestations, will have high titers of trophozoite-specific antibody, whereas *E. dispar*-infected have none[104,108,156]. However, serum antibody responses will remain long after cure, preventing discrimination of past and present infections[157].

Several *Entamoeba* antigen detection ELISAs and rapid strip tests have been developed and are commercially available (see Chapters II and IV), but some are unable to distinguish *E. histolytica* and *E. dispar*[150]. All of the tests are limited to unfixed trophozoites, which are not predominant nor stable in stool[105,150,158]. A larger problem is that these tests have sensitivities that are 1000-10000 times less than conventional PCR assays[159,160]. Quantitative PCR (qPCR) assays have recently come to the forefront as *Entamoeba* diagnostic gold standards[151,153,161]. However, PCR and qPCR assays can involve laborious DNA extraction methods, and, being more expensive than antigen detection, they are not accessible to laboratories where *E. histolytica* is endemic[150]. Thus, there is still promise and benefit to pursuing new biomarkers for *Entamoeba* detection, particularly those specific to pathogen cysts shed in stool. The second goal of this dissertation work is to validate a novel cyst biomarker in stool that could be the target of more sensitive *E. histolytica* detection assays.

III. DISSERTATION AIMS AND CONTRIBUTIONS

This dissertation was conducted under the Accelerated Molecular Probe Pipeline (AMPP) research consortium and had two aims. The first aim was to develop a new accelerated affinity probe pipeline, which would employ previously established yeast display of recombinant antibodies while overcoming the limitations of obtaining soluble antibody needed for diagnostic applications. The second aim was to investigate and validate a novel *E. histolytica* cyst biomarker, ideally with the affinity reagents obtained from the accelerated pipeline.

As described Chapter 2, I began these endeavors with rational selection and recombinant production of *E. histolytica* candidate biomarkers. I utilized recent cyst proteomic and transcriptomic data, as well as bioinformatics programs, to identify antigens and regions within antigens that could be highly sensitive and specific markers of *E. histolytica* infection. As *Entamoeba* cysts cannot be obtained in large quantities *in vitro*, bacterial recombinant expression was undertaken, with assistance from the Seattle Structural Genomics Center for Infectious Disease. Primarily using medium- and high-throughput approaches and some eukaryotic expression, eighteen (37%) of the chosen candidate biomarkers were obtained as soluble antigen.

Next, as detailed in Chapter 3, I employed yeast display to select scFv antibody fragments specific to *E. histolytica* candidate biomarkers. A total of 34 yeast-displayed scFv (yeast-scFv) clones detecting seven antigens were successfully obtained within two to three weeks of work. In conjunction with Maria Kahn at PATH, a lyophilization protocol that can preserve yeast-scFv for at least one year up to 45°C was developed. From there, utilizing lyophilized yeast-scFv stocks, our collaborators at the University of Queensland and I generated crude, scFv-displaying yeast cell wall fragments and investigated their functionality in a number of assays. The fragments, named nanoyeast-scFv, demonstrated sensitive (pg/mL) and specific

antigen detection in a complex stool matrix when tethered onto screen-printed gold electrodes. The nanoyeast-scFv also bound antigen specifically and down to 1 pg/mL in a multiplex, surface-enhanced Raman spectrometry (SERS) assay. Finally, though having a higher limit of antigen detection than the electrochemical and SERS assays, a nanoyeast-scFv ELISA assay again demonstrated tethering and enrichment of working yeast-scFv fragments onto a solid surface.

Finally, in Chapter 4, I characterized traditional murine monoclonal antibodies generated specifically against a divergent and polymorphic region of the Jacob2 cell wall protein. I was successful in identifying antibodies with apparent specificity to *E. histolytica* recombinant protein. Two of the antibodies bound fixed, cyst-like objects in microscopy smears of xenic *Entamoeba* cultures, with one of the antibodies demonstrating specificity to pathogenic *E. histolytica* parasites as compared to Entamoeba commensals. These results showed promise in providing affinity reagents functional with fixed stool samples and in validating a novel *E. histolytica* cyst biomarker for improved detection.

CHAPTER II: RATIONAL SELECTION AND RECOMBINANT EXPRESSION OF *ENTAMOEBA HISTOLYTICA* CANDIDATE BIOMARKERS

Introduction

Biomarkers are indicators of biological conditions, such as a human illness[162–164]. In infection, this is often a component of pathogen found reproducibly in host tissues and fluids. Biomarkers are important to the development of accurate diagnostic tests[162–164]. A good biomarker is sensitive, detectable in a large proportion of diseased cases, and specific, i.e. able to distinguish the causative agent from other present microbes.

Several *Entamoeba* biomarkers have been investigated prior to this project. *Entamoeba* molecules producing species-specific monoclonal antibodies include peroxiredoxin[165–168], the intermediate and heavy subunits of the Gal/GalNAc lectin[169–171], *E. histolytica* lipophosphoglycan[159,172], a 30 kDa *E. dispar* hypothetical protein[173], and an unknown cyst antigen[174]. The peroxiredoxin has become the target of a commercial assay, the Triage Parasite Panel, but the assay is unable to distinguish between *E. histolytica* and *E. dispar*[150,175–177]. Antibodies against another biomarker, the serine-rich *E. histolytica* protein (SRHEP), have been incorporated into the *E. histolytica*-specific Optimum S kit, but weak cross-reaction was observed when sample containing $\geq 10^4$ *E. dispar* trophozoites was assayed[159,178,179]. To date, the most developed and successfully detected biomarker is the heavy subunit of the Gal/GalNAc lectin, a virulence factor expressed on the surface of *E. histolytica* trophozoites[160,171,180–182]. This protein subunit has sequence differences between *E. histolytica* and *E. dispar*, while remaining

conserved amongst *E. histolytica* strains[183,184]. In studies to date, detection assays of this target were found to be 79% sensitive and 96% specific to gold-standard nucleic acid detection[161].

The reported gap in the lectin's diagnostic sensitivity may have an impact on *E. histolytica* diagnosis, with reports of missed cases and low sensitivities in some epidemiological studies[158,160,185]. It has been speculated that a key problem rests on the assays' inability to detect cysts[158]. However, detection utilizing cyst stage biomarkers has been largely uninvestigated, despite the fact that cysts are more abundant and stable than trophozoites in stool[111,114]. Here, with the hope of identifying a novel biomarker for *E. histolytica* detection, I aim to rationally select candidate cyst biomarkers to which antibody detection reagents will be generated.

Results

Candidate Biomarker Selection

Selection of about 50 candidate biomarkers was conducted using recent proteomic and transcriptomic datasets and applying rational filtering procedures. This task is summarized in **Figure 2-1**. The AMPP research consortium previously collected five *E. histolytica*-positive, semi-purified stool samples, lysed the parasites and other cells by freeze-thawing or sonication, and subjected the lysates to liquid chromatography tandem-mass spectrometry (LC-MS/MS)[186]. This was the first proteomic study investigating native *E. histolytica* cysts. Although the proportion of peptides in the samples that were uniquely identified as *E. histolytica* were small (6.9-10.2%), a list of 417 candidate biomarkers were generated[186]. One hundred and ninety-five of those biomarkers had not been previously identified in trophozoite proteomic and EST data, including four cyst wall proteins[186].

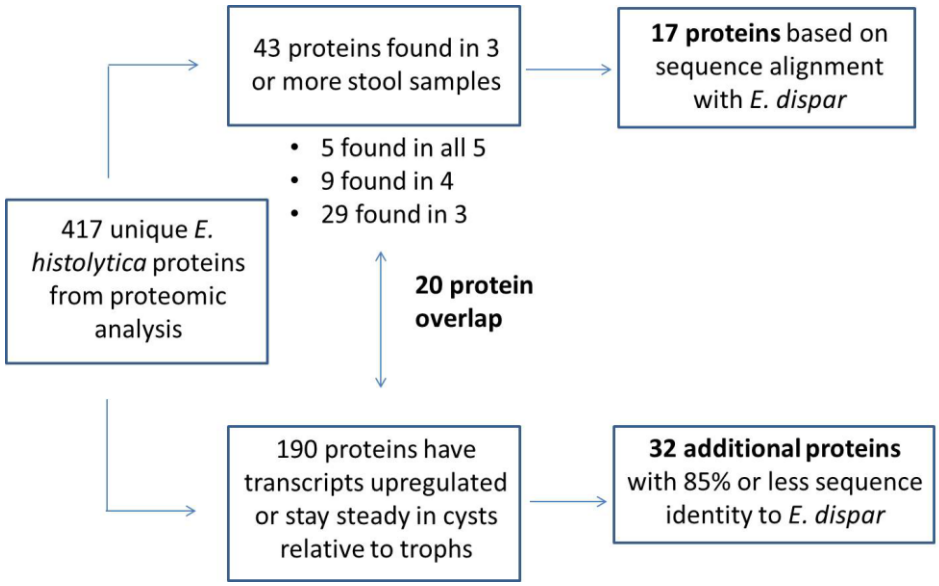

Figure 2-1. A schematic representation of the AMPP selection strategy of *E. histolytica* candidate biomarkers. The original list of 417 proteins was generated in a mass spectrometry investigation of *E. histolytica* cysts obtained from clinical stool. Subsequent reduction of the list was based on presence of biomarkers across a majority of samples (top arm)[186] and on the biomarkers' mRNA transcript regulation in the cyst stage (bottom arm)[138].

Limiting the list to samples found in three or more of the stool samples generated a list of 43 candidates, with five proteins appearing in all five tested stool samples, nine in four out of five samples, and 29 in three out of five samples (**top arm of Figure 2-1**). Median percent identities with *E. dispar* homologues was 98% (range 93-99%), 96% (range 82-99%), and 95% (range 36-99%) for targets in 5, 4, and 3 out of 5 stool samples respectively (**Table 2S1-Excel File**). Based on individual sequence alignments with their reported *E. dispar* homologues, seventeen targets had localized, divergent domains or regions and were selected for further studies. When possible, only these divergent regions were chosen for cloning and expression, to decrease the likelihood of obtaining cross-reactive reagents, and predicted signal sequences and transmembrane domains were not included. One potential target, ubiquitin, was eliminated due to its small molecular weight (<20 kDa). Biomarker EHI_006810 was accidentally included on this list, due to using an incorrect *E. dispar* protein as its closest homologue.

Returning to the complete biomarker list, candidates were also sorted by their reported fold change in detected mRNA between spontaneously encysting xenic cultures and cyst-less axenic laboratory strains (**bottom arm of Figure 2-1**)[138]. Regardless of p-value, there were 190 proteins whose mRNA transcript levels increased or did not change in encysting cultures relative to axenic cultures (fold changes of ≥1) (**Table 2S2-Excel File**). Twenty of them previously appeared in selection lists for being present in ≥3 of five stool samples, of which nine were already marked for expression and affinity reagent generation. Setting ≤85% sequence identity with *E. dispar* homologues as a second cutoff, I identified thirty-two additional proteins with this strategy and added them to the list of targets to be expressed.

A small number of proteins were added to the pipeline outside of the described filters (**Table 2-1**). Prior to the mass spectrometry study, EHI_115350, EHI_000780, EHI_142000, and EHI_044500 were selected as targets for their upregulated mRNA during encystation and were successfully expressed, but they ultimately appeared in 0-2 stool samples by mass spectrometry. EHI_182030, also appearing in only one stool sample, was accepted after successful bacterial expression by the SSGCID. Adding these targets to the 49 rationally selected targets in Figure 2-1, the total number of selected candidate biomarker is 54.

TABLE 2-1. A Subset of *Entamoeba histolytica* Proteins Selected for Investigation in Spite of Rational Filters.

E. histolytica Locus ID	Annotation	Found in Stool Samples (Out of 5)[a]	Transcriptome data[b]		*E. dispar* homolog		
			fold change: cyst / troph	p-value	Locus ID	% identity	% *E. his* coverage
EHI_044500	Cyst wall-specific glycoprotein Jacob	0	0.0025	1290.32	EDI_246160	64	100
EHI_142000	histone acetyltransferase, putative	0	7.8E-07	5.78	EDI_338230	92	99
EHI_115350	chromodomain-helicase-DNA-binding protein, putative	0	0.0083	3.08	EDI_136060	98	100
EHI_000780	chromodomain-helicase-DNA-binding protein, putative	2	0.0006	2.28	EDI_026450	94	100
EHI_182030	Rab Family GTPase	1	0.1280	-1.70	EDI_154730	48	100

[a]From source [169]
[b]From source [170]

Bacterial Expression of Entamoeba histolytica *Targets*

Bacterial expression was attempted for 45 out of the 49 rationally selected targets, primarily utilizing 4 different bacterial vectors. Gene synthesis, commercial cloning services, and a medium-throughput, small-scale protein expression screen developed by the SSGCID (**Figure**

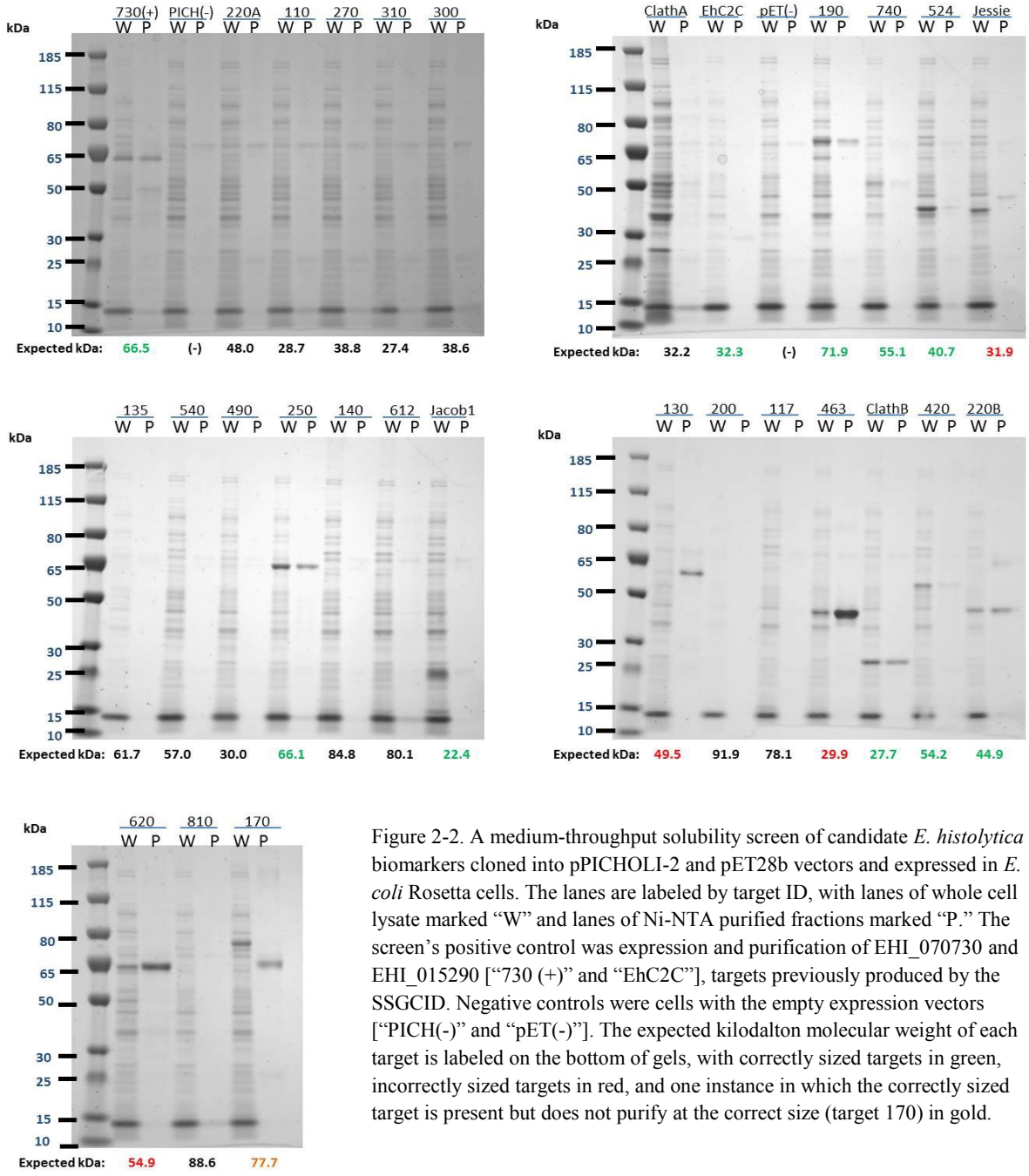

Figure 2-2. A medium-throughput solubility screen of candidate *E. histolytica* biomarkers cloned into pPICHOLI-2 and pET28b vectors and expressed in *E. coli* Rosetta cells. The lanes are labeled by target ID, with lanes of whole cell lysate marked "W" and lanes of Ni-NTA purified fractions marked "P." The screen's positive control was expression and purification of EHI_070730 and EHI_015290 ["730 (+)" and "EhC2C"], targets previously produced by the SSGCID. Negative controls were cells with the empty expression vectors ["PICH(-)" and "pET(-)"]. The expected kilodalton molecular weight of each target is labeled on the bottom of gels, with correctly sized targets in green, incorrectly sized targets in red, and one instance in which the correctly sized target is present but does not purify at the correct size (target 170) in gold.

2-2) allowed me to more rapidly determine our targets' solubility and to focus large-scale efforts on high yield recombinants. Cloning and expression success of targets by vectors is summarized in **Table 2-2**. The fourth vector (not shown) was a bacterial and yeast shuttle vector called pPICHOLI-2 (Mobitec, Göttingen, Germany), which expressed none of the 20 targets cloned into it (**Table 2S3**). Additionally, these results do not include EHI_070730, successfully purified with an MBP tag by the SSGCID, but do include EHI_015290, which was also expressed as a C-terminal fragment with a SUMO tag by University of Virginia. Depending on the vector utilized, 10-53% of cloned targets failed to express, and 29-90% of expressing targets failed to be soluble (**Table 2-2**).

Ultimately, soluble expression in *E. coli* systems was achieved in 16/45 (35.6%) of targets overall (**Tables 2S4, 2S5, 2S6**). Eight biomarkers were tested with the pET27b periplasmic expression vector, and, while five were successfully cloned, three (37.5%) resulted in soluble, tandem mass spectrometry-confirmed protein (**Table 2S4**). When visualized on SDS-PAGE, only one of those targets, EHI_101240, was non-degraded and pure, but its band was a higher molecular weight than anticipated. Greater success was seen when 23 targets were gene synthesized and cloned into pET28b (**Table 2S5**). After high-throughput IMAC purification, eight targets were visibly soluble and were confirmed by mass spectrometry, though three targets and one target had higher and lower molecular weights than expected, respectively. Another

TABLE 2-2. Cloning and Expression Success Rates of Candidate *E. histolytica* Biomarkers, by Bacterial Expression Vectors.

Vector	N-terminal Tags	C-terminal Tags	Biomarkers Attempted, n		Cloned, n (%)		Expressed, n (%)		Soluble, n (%)		
			Whole	Fragment	Whole	Fragment	Whole	Fragment	Whole	Fragment	Total
pET27b	pelB leader sequence	6xHis, FLAG	0	8	--	5 (62.5%)	--	ND	--	3 (37.5%)	3 (37.5%)
pPICHOLI-2	6xHis, BCCP*	BCCP, FLAG	10	10	10 (100%)	10 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
pET28b	6xHis, T7	BCCP, FLAG	15	14	15 (100%)	14 (100%)	7(46.7%)	7 (50.0%)	5 (33.3%)	4 (28.5%)	9 (31.0%)
SSGCID (pET14)	6xHis	None	27	11	23 (85.2%)	10 (90.9%)	15 (55.5%)	9 (81.8%)	9 (33.3%)	1 (9.1%)	10 (26.3%)

ND= not determined

target, EHI_145840, expressed soluble and correctly-sized protein after its gene was removed from pPICHOLI-2 and placed into pET28b.

Finally, the SSGCID took on the expression of a large proportion (38/45) of the attempted targets (**Table 2S6**). Five were not cloned, either due to failing SSGCID selection filters (i.e. too many cysteines, transmembrane domains) or to PCR failure. Targets that were more than 100 kDa molecular weight were expressed in 2-6 fragments, but almost all of these fragments were insoluble upon expression. Overall, 26% of targets were expressed in soluble form, and there was a trend of proteins expressing at a higher than expected molecular weight, as seen with the previous expression screens.

Large Scale Expression and Purification

Table 2-3 lists the *E. histolytica* targets that were produced large-scale for affinity reagent selections. Including the special targets listed in **Table 2-1** and an additional target produced in a cell-free system (described below), there were 23 targets available. Overall, most targets that were soluble in small-scale screens were soluble when generated in large cultures. Inclusion body purification and refolding was attempted on three targets and greatly added to the yield of two of proteins (EHI_101240 in pET27b, EHI_070730 from SSGCID). The largest problem encountered in obtaining the required yield for necessary experiments was the presence of co-purified *E. coli* contaminants, which was often corrected with the use of TALON resin. Additionally, a few proteins, such as the two C2 domain-containing proteins (EHI_118130 and EHI_015290) would be expressed in large amounts but would often precipitate out of solution post-purification.

TABLE 2-3. The *Entamoeba histolytica* Recombinant Biomarkers Produced in Large-scale Cultures.

Target Locus	Annotation	Whole Sequence?	Residues Expressed	Expected Molecular Weight (kDa)[b]
EHI_104630	Filamin 2	No	569-769	26.0 (pET27b)
		No	569-769	29.9 (pET28b)
		Yes	1-857	96.4
200.m00090	Clathrin heavy chain, putative	No	940-1116	27.7
		No	1-319	36.0
		No	400-736	39.0
		No	1114-1448	39.0
EHI_006810	14-3-3 protein 3	Yes	1-240	27.5
EHI_015290[a]	C2 domain protein, putative	No	143-281	27.5
		Yes	1-281	30.0
EHI_033250	Polyadenylate-binding protein, putative	No	1-205	23.2
		Yes	1-521	66.1 (pET28b)
		Yes	1-521	59.8 (SSGCID)
EHI_096190	Hypothetical protein 6.t00015	Yes	1-549	71.9
EHI_101240	Nucleic acid binding related	No	1-192	21.5
		Yes	1-843	95.6
EHI_109890	Chitinase, putative	Yes	16-577	58.0
EHI_110170	Ubiquitinyl hydrolase, putative	No	437-1036	77.7
EHI_114220	C2 domain containing protein, putative	No	1903-2227	44.9
EHI_118130[a]	C2 domain protein, putative	Yes	1-389	49.5
EHI_136160	Calreticulin, putative	Yes	14-389	45.1
EHI_145840	Peroxiredoxin, putative	Yes	1-233	33.2
EHI_158620	Opioid growth factor receptor (OGFr) conserved region	Yes	1-402	54.9
EHI_166420	Cyclin, putative	Yes	1-400	47.9
EHI_070730	Rho GTPase	Yes	1-194	66.5
EHI_138010	Hypothetical protein	Yes	1-421	48.0
EHI_192430	Hypothetical protein	Yes	1-251	29.6

[a]Unstable when produced large-scale.

[b]Including tags.

Expression and Purification of Targets in Pichia *and Cell Free Systems*

Two targets (EHI _104230 and EHI_138010) were attempted in the SSGCID's *Pichia pastoris* expression system. Only EHI_104320 had visibly expressed protein secreted into culture medium. However, when the medium was subjected to IMAC purification, the protein did not bind the column and hence was not purified. Meanwhile, five targets, including EHI_138010, were tested for expression in the SSGCID's wheat germ cell free system[187]. Targets EHI_152170 and EHI_065240 did not amplify with their designed PCR primers (**Table 2-4**), while targets EHI_064490 and EHI_010740 did not express soluble protein in small-scale reactions. However,

EHI_138010 produced 0.2-0.4 mg/mL protein in a large scale translation reaction. Overall, this is a success rate of 20% for the WGCF system.

Discussion

As purified *E. histolytica* cysts cannot be obtained in large quantities, production of affinity reagents against them must be conducted on individual, pre-determined targets[136]. To this end, a rational selection of candidate *E. histolytica* cyst biomarkers in stool was undertaken by filtering cyst-derived proteomic and transcriptomic datasets on features that could be characteristic of a sensitive and specific biomarker. The cyst proteomic dataset was generated by our research consortium, with validation of a novel cyst biomarker as a goal[186]. That investigation was the first to characterize native *E. histolytica* cysts by mass spectrometry, and it identified 417 *E. histolytica* proteins, 195 of which had never been identified by studies in trophozoites and 4 of which were already known as cyst cell wall proteins.

A two-pronged filtering scheme was applied to choose candidate *E. histolytica* biomarkers. The first prong narrowed the list down to targets that were in ≥3 or more stool samples in proteomic analysis. Though this method generated 43 candidate proteins, only 4 of them had <90% sequence identity with an *E. dispar* protein and were larger than our minimum size of 15 kDa. Two of those hits, EHI_109890 and EHI_152170, are known components of the *E. histolytica* cell wall[143,144]. By selecting to express only localized regions or domains within targets that have low sequence identity with *E. dispar* homologues, the list was ultimately boosted to 17 targets. The second arm of the selection scheme identified proteins that had upregulated or maintained mRNA levels in cyst-containing cultures relative to laboratory strains[138]. One hundred ninety proteins in the original list fell into this category, of which 22

were also in ≥3 stool samples. From among these 190 proteins, thirty-one additional markers with ≤85% sequence identity with *E. dispar* were chosen. The Gal/GalNAc lectin heavy subunit, a target of *E. histolytica* diagnostic tests, is 86% conserved between *E. histolytica* and *E. dispar*, and, hence, 85% identity for a cut-off seemed appropriate[183]. The final list had 54 biomarkers, including five special targets accepted outside of the selection scheme, nine proteins falling into both prongs used for selection, and thirty biomarkers that were never identified in trophozoite proteomes and EST databases[186].

Some of the potential biomarkers identified have also been found and noted in other *Entamoeba* studies. A recent trophozoite cell surface proteome also identified 11 of our targets, nine of which were in ≥3 stool samples[188]. Eight targets, most again falling in the abundant category, had mRNA transcripts upregulated two-fold or more in the virulent HM1:IMSS strain versus the avirulent Rahman strain[189]. Among smaller studies on this organism, peroxiredoxin (EHI_145840) has been established as an important, differentially expressed virulence factor, the heat shock 90 protein (EHI_102270) has been identified as a potential drug target, and the calreticulin molecule (EHI_136160) has been found to be very immunogenic and important in phagocytosis[127,190–192].

There are some limitations to the methods of this biomarker selection. First, it is possible that proteins of interest to this study were missed in proteomic analyses. One such missed target was the Jacob2 lectin: this abundant, outer wall protein confirmed the presence of cysts in immunofluorescent smears of the studied stool samples, but it was not identified even once in the same stool samples when they were subjected to mass spectrometry[145,186]. Proteins such as Jacob2 could have degraded and disappeared entirely before detection, or their interactions with cell wall components could have rendered them insoluble and unextractable. Jacob2 also has a

low complexity domain, which would have been difficult to detect.[193,194] Additionally, the high background (~90%) created by human, plant, and bacterial proteins could have masked some proteins, though it is likely that those missed were less abundant relative to those that were identified in these complex mixtures. Finally, some targets perhaps could not be cleaved effectively by the employed protease or ionized by the mass spectrometer, and they would have been missed entirely[194]. However, with the limitations of mass spectrometry in mind and with the hope of developing a robust candidate list, I also included additional targets based on a cyst transcriptome[138].

Recombinant expression was undertaken, as *E. histolytica* cannot be efficiently encysted *in vitro*[136], and it was a challenge. Spreading cloning and expression screening of the targets across multiple research groups, multiple expression systems, and multiple vectors, we found 18 proteins (36.7%) on my selected list to be soluble in small-scale expression screening. In a high proportion of cases, these targets were also obtained as soluble product in large scale culture, but some proteins (i.e. C2 domain-containing proteins) were unstable in solution. Others misfolded and accumulated in inclusion bodies, and, though some yields were improved by inclusion body extraction and refolding; this process required additional time, labor, and care. Many soluble proteins were higher molecular weights than anticipated on SDS-PAGE (typically by 5-12 kilodaltons); it could be the result of secondary structure and incomplete denaturation, but, as most targets were confirmed by mass spectrometry, the issue has not been further investigated.

Most of the recombinant expression strategies employed resulted in similar rates of success (20-37.5%). Not including the *Pichia pastoris* system, where only two targets were tried, the only expression system with consistent failure was bacterial expression with the pPICHOLI-2 bacteria/yeast shuttle vector. This could have been due to incompatibility of the expression

system with autoinduction media, the high temperature sensitivity of the selection drug, mixed bacteria/yeast codon optimization, or the ORF's numerous epitope tags. Factors that could have contributed to failed expression in the other systems could have included location of epitope tags, product degradation, codon usage, and lack of proper folding machinery. In future endeavors, additional vectors and expression systems, especially those compatible with eukaryotic expression, could be considered. To facilitate quick recloning and investigations in a wider range of systems, ligation independent cloning or the Gateway vector system could be considered. Additionally, synthesis of peptides could have been a quick way to obtain antigen, specifically for targets with short divergent regions or gaps; affinity reagent selections against linear peptides tends to be less fruitful but is not impossible.

In summation, out of list of more than 400 candidate *E. histolytica* cyst biomarkers, around fifty were identified with expression patterns conducive to accurate pathogen diagnosis in stool, and eighteen could be expressed heterologously in a soluble form.

Methods

Candidate Biomarker Selection

A list of candidate *E. histolytica* biomarkers was generated by proteomic analysis as described in Ali et al., PLoS NTDs, 2012[186]. *E. histolytica* cysts were semi-purified from five stool samples obtained in Dhaka, Bangladesh. The cysts were sonicated, freeze-thawed or both to obtain a pool of proteins that were subjected to liquid chromatography tandem mass spectrometry. A protein was placed on the list if its detected peptide sequence was at least partially unique to *E. histolytica*.

A series of filters were applied to the mass spectrometry list to narrow down candidate biomarkers. First, the list was filtered to targets which appeared in 3 or more stool samples (out of n=5). After that, top targets still remaining were ClustalW aligned with reported *E. dispar* homologues using BioEdit v.7.0.9.[195], and localized divergent regions in the alignment were noted. The percent identity of whole targets and fragments were obtained via Blastp or Geneious. In a parallel investigation, the list was sorted in descending order by the fold difference in RNA transcript as detected in a spontaneously encysting *E. histolytica* xenic culture versus established, non-encysting laboratory strains[138]. Targets less than 15 kDa size were not accepted. Retrospective BLAST analyses of selected targets against *E. dispar* and mammalian targets were performed using Geneious version 6.0.3. Signal peptides and transmembrane domains were identified by SignalP3.0 and TMHMM Server v.2.0[196].

Cloning of E. histolytica *Target Sequences into Bacterial Expression Vectors*

Targets EHI_115350 (350), EHI_00780 (780), EHI_142000 (142), EHI_015290 (EhC2C), and EHI_044500 (Jacob) were generated and provided by the University of Virginia prior to bioinformatics analysis[197]. 350, 780, and 142 were PCR-amplified and cloned into a modified pET11a vector with an N-terminal 6xHis tag (Novagen, Merck, Darmstadt, Germany). EhC2C and Jacob were codon optimized and synthesized by GeneArt and placed in a pET-SUMO vector with an N-terminal 6xHis-SUMO tag (Life Technologies, Carlsbad, CA). All targets were expressed in BL21 (DE3) *E. coli* cells.

Many of the candidate biomarkers were placed in the expression pipeline of the Seattle Structural Genomics Center for Infectious Disease (SSGCID). Most proceeded through the pipeline's first tier, which utilizes pET14b-derived vectors with N-terminal His tags[198]. One

target, EHI_070730, was cloned into a standard vector modified with an N-terminal maltose-binding protein (MBP) tag[199]. All targets were expressed in BL21 Rosetta (DE3)R3 *E. coli* cells.

An initial batch of the candidate biomarkers was cloned into a modified pET27b vector (Novagen) in which the HSV tag had been switched out for a FLAG tag. Coding sequences were isolated from *E. histolytica* strain HM1:IMSS genomic DNA by PCR, with 50 µL reactions consisting of 1x Phusion High-Fidelity Mastermix (New England Biolabs, Inc., Ipswich, MA), 2.5 ng/µL genomic DNA, and 0.2 µM forward and reverse primers. The amplification program consisted of 1 denaturation cycle 95°C for 30s; 35 amplification cycles of 95°C for 10s, 60°C for 30s, 72°C for 30s; and 1 extension cycle at 72°C for 5 minutes. The primers used were generated by Primer3.0 and are listed in **Table 2-4**. The primer combinations conferred a 5' NcoI or BsmBI and a 3' NotI restriction site to amplified products. Successfully amplified products were purified with a QIAquick PCR purification kit (Qiagen, Hilden, Germany), and 0.5 µg of each insert was double-digested for 1 hour at 37°C with 5 units NcoI and NotI enzyme or for 1 hour at 37°C and 1 hour at 55°C with 5 units BsmB1 and NotI enzyme (New England Biolabs). Next, after another PCR purification, 50 ng of doubly digested inserts were ligated into 50 ng of NcoI- and NotI-digested pET27b modified vector. Resultant ligations were isopropanol precipitated and electroporated into Alpha-select *Escherichia coli* cells (Bioline, London, UK). Correct ligation and orientation was confirmed via colony PCR with pelB primer or inserts' forward primers paired with inserts' reverse primers (**Table 2-4**), again utilizing Phusion High-Fidelity Mastermix and the PCR program described above. Gel electrophoresis of NcoI/NotI doubly digested plasmids and sequencing were also employed to confirm ligation.

To produce remaining targets that were not already successfully expressed or obtained from the SSGCID, genes were codon optimized and synthesized by Genscript (Piscataway, NJ).

TABLE 2-4. The Primers Utilized in Cloning of Recombinant *E. histolytica* Biomarkers.

Primers for *E. histolytica* gene amplication- pET27b cloning	
Primer	**Sequence**
EHI_136160 Forward	ATTTCCATGGATGAAGATTGGGATATGTTAGC
EHI_136160 Reverse	ATAAGCGGCCGCAAGCTCTTCTTTGTTTTCTTCTTTC
EHI_102270 Forward	AAAACCATGGGAAATAGAAAAGATCAATCATTAG
EHI_102270 Reverse	ATTTGCGGCCGCTTGTTCCCATTCATGTGTAACTTTC
EHI_104630 Forward	CAGCCCATGGTTGTTGATGGAGAAGGTATTG
EHI_104630 Reverse	TTTAGCGGCCGCATCTGCATCAGCATCTTCTTC
EHI_060680 Forward	ATTTCCATGGGTGTTGCAGAACTTTGTATTG
EHI_060680 Reverse	AAAAGCGGCCGCGAACTTATTGACTTGGAATTTAACTG
EHI_101240 Forward[a]	AAAACGTCTCACATGCCAGCACAAGCTAATAAGAAG
EHI_101240 Reverse	ATTTGCGGCCGCTTGTCCTGGTGCATCAACTTTC
EHI_012500 Forward	ATTTCCATGGTGATGTATAATATTTGTTGTGTATCAACT
EHI_012500 Reverse	ATTAGCGGCCGCATAAAGACTTTGTGATAGTTCCAGTG
EHI_033250 Forward[a]	ATTTCGTCTCACATGACCACTGCTCAAAGTAATTTTTC
EHI_033250 Reverse	AAAAGCGGCCGCGAATTGTTCTTCAGTAACAGTAGCATC
EHI_114220 Forward	AAAACCATGGGAAAAAGTCTTGAACCAATGAAAG
EHI_114220 Reverse	AAAAGCGGCCGCAATTGCAGTGATAACAGAAACTGAAG
[a]Confers 5' BsmB1 site instead of a NcoI site	
Primers for colony confirmatory PCR	
Primer	**Sequence**
pelB	GCTGCTGGTCTGCTGCTC
T7 promoter	TAATACGACTCACTATAGGG
T7 terminator	GCTAGTTATTGCTCAGCGG
Primers for *E. histolytica* gene amplification- Wheat Germ LIC Cloning	
Primer	**Sequence**
490 Forward	ACCACCACCACCATATGGAATTTCTGAATATCAATGGTTCGTGCGT
490 Reverse	CCCTATATATGGA TCCTCACTTATCGTCATCGTCTTTGTAGTCG
740 Forward	ACCACCACCACCATATGAAGGAAGTGAAAGAACGCAACG
740 Reverse	CCCTATATATGGA TCCTCACTTATCGTCATCGTCTTTGTAGTCG
Jessie Forward	ACCACCACCACCATATGCTGAACATTACCTTCTCGCAGCGCA
Jessie Reverse	CCCTATATATGGATCCTCACTTATCGTCATCGTCTTTGTAGTCGG
524 Forward	ACCACCACCACCATATGGAAGACAACATTAAGAAGGGCATCCTGG
524 Reverse	CCCTATATATGGA TCCTCACTTATCGTCATCGTCTTTGTAGTCGG

Twenty targets were codon optimized for both *Pichia pastoris* and *Escherichia coli*, double digested via added 5' SalI and 3' NotI restriction sites, and ligated into the pPICHOLI-2 shuttle vector (MoBiTec, Göttingen, Germany), while 23 targets were codon optimized for *Escherichia coli*, digested via added 5' SalI and 3'NotI sites, and cloned into the pET28b vector (Novagen).

Each synthesized gene also had a biotinylation (BCCP) site, a FLAG tag, and two stop codons added to the C-terminus before the NotI site. The decision to add a second, C-terminal BCCP site for targets cloned in pPICHOLI-2, which is supposed to have an N-terminal biotinylation site, was made when noticing the N-terminal site was missing important amino acids on both ends[200]

The requested pET28b clones were confirmed by colony PCR with T7 promotor and terminator primers (**Table 2-4**). This endeavor was first tried with Phusion HF mastermix reaction, as described above. The PCR program consisted of 1) denaturation at 95°C for 30 seconds; 2) amplification for 30 cycles at 95°C for 10 seconds, 46°C for 30 seconds, and 72°C for 15 seconds; and 3) extension for 72°C for 5 minutes. In a second attempt, reactions with Amplitaq polymerase were prepared, consisting of 1x PCR buffer with magnesium chloride, 200 µM dNTPs, 0.5 µM primers, and 1.25 units Amplitaq DNA polymerase. The reactions were subjected to 1) denaturation at 94°C for 2 minutes; 2) 30 cycles of 94°C for 1 minute, 50°C for 1 minute, and 72°C for 2 minutes; and 3) final extension at 72°C for 5 minutes.

Six genes in pPICHOLI-2 were later double digested with SalI and NotI and re-ligated into the pET28b vector.

Expression Studies

Excepting special and SSGCID targets, all confirmed vectors were purified via standard mini-prep protocol (Qiagen) and electroporated into BL21 Rosetta 2(DE3) cells (EMD Millipore, Billerica, MA) for expression. Small scale cultures of targets cloned into pET27b were induced with 0.25 mM IPTG (EMD Millipore) at exponential growth and were shaken at 250 rpm overnight at 20°C. The next day, cells were pelleted by centrifugation for 10 minutes at 10,000 x g and 4°C. Osmotic shock was employed to extract protein, first with a 10 minute incubation of

the cell pellet in osmotic shock buffer (20 mM HEPES, pH 8.0, 1 mM EDTA, 20% w/v sucrose) on ice, followed by another spin. Supernatants were saved, and the pellet incubated with ice cold 5 mM magnesium sulfate for 15 minutes, followed by another spin and collection of supernatant. The combined supernatants obtained from this extraction (~50 mL/100 mL culture) were mixed with 1 mL Ni-NTA resin (Qiagen) overnight at 4°C. Next, the resin was centrifuged for 5 minutes at 931 x g, and the resulting resin pack was mixed with 50 mL TALON wash buffer (50 mM sodium phosphate, pH 7.0, 300 mM sodium chloride, 20 mM imidazole) for 10 minutes and centrifuged again. The resin was then placed into a column, where it was washed with a minimum 20 mL of additional wash buffer and until no protein was detected in the flow-through by Bradford assay. Finally, protein was eluted from the resin with TALON elution buffer (50 mM sodium phosphate, pH 7.0, 500 mM sodium chloride, 300 mM imidazole) until it was no longer detected by Bradford. The resultant eluates were subjected to 4-12% Bis-Tris SDS-PAGE, transferred to a PVDF membrane, blocked with StartingBlock buffer (Life Technologies), and probed with 1 μg/mL α-FLAG antibody (#F1804, Sigma Aldrich, St. Louis, MO) and 1:5000 HRP-conjugated anti-mouse IgG (H+L) secondary (#1706516, Bio-Rad, Hercules, CA) . Purified protein bands were extracted from SDS-PAGE and analyzed by tandem mass spectrometry at the Fred Hutchinson Cancer Research Center (Seattle, WA).

The expression of pET28b targets was screened utilizing high throughput IMAC purification as described by the SSGCID[201]. For this screen, previously characterized SSGCID clones expressing EHI_070730 and EHI_015290 were utilized as positive controls, whereas a clone expressing empty pET28b vector was employed to establish background protein expression. Autoinduction of pPICHOLI-2 clones was concurrently attempted, with incorporation of a low salt ZY (Luria Broth) base.

The prepared resultant whole cell lysate and Ni-NTA-purified fractions were separated on 4-12% Bis-Tris SDS-PAGE. Purified fractions of interest were also re-run as a Western Blot and probed with 1 μg/mL α-FLAG antibody and 1:5000 goat anti-mouse-HRP. Most soluble targets were excised from SDS-PAGE and analyzed by LC-MS/MS at MS Bioworks (Ann Arbor, MI).

Large Scale Expression

The pET27b set of targets were expressed in 2 liter batches, with induction of 0.5 mM IPTG and subsequent shaking conducted at 20-25°C. Osmotic shock extraction and Ni-NTA IMAC conducted as described above.

The SSGCID Tier 1 autoinduction and purification protocol was utilized for upscaling successfully expressing pET28b targets, as well as the special targets obtained from the University of Virginia. This protocol is very similar to the small-screen described above. Two to four x 0.5 liters of autoinduction media were inoculated with *E. coli* starter culture. After 72 hours, *E. coli* were harvested, resuspended with SSGCID lysis buffer, and sonicated (3-5 x 5-30 second pulses). With addition of benzonase (EMD Millipore), the lysate was rocked for 40 minutes at room temperature, centrifuged at 10,000 x g for 1 hour, and filtered through a 0.45 μm cellulose acetate filter (VWR International, Radnor, PA). Protein was purified from lysates with Nickel-NTA utilizing the SSGCID wash and elution buffers. For some targets that were noted to have *E. coli* protein contaminants, purification was undertaken with TALON resin and buffers (Clontech Laboratories, Inc., Mountain View, CA).

To obtain proteins contained in inclusion bodies, expressing *E. coli* cells were lysed, either by 1 hour incubation with Bugbuster lysis buffer (EMD Millipore) supplemented with

lysozyme (Sigma Aldrich) and benzonase or by sonication. The lysate was centrifuged for 5-10 minutes at 500 x g, and the resultant pellet was mixed with 50 mL 100 mM Tris buffer, pH 8.0 with 4-6 molar urea overnight at 4°C. The mixture was then centrifuged for 5-10 minutes at 10,000 x g, and the supernatant was mixed with Ni-NTA resin equilibrated with the urea buffer. Purification was performed with typical wash and elution buffers supplemented with 4-6 M urea. Eluates were pooled and placed into 3.5 kDa cellulose tubing (Fisher Scientific International, Hampton, New Hampshire) or 10 kDa Slide-a-lyzers (Life Technologies) and dialyzed 2-3 days at 4°C, each day getting transferred into a buffer with 2 M less urea. Over the last 24 hours of dialysis, the eluates were dialyzed into a phosphate buffer. Any precipitating material was centrifuged at high speed and discarded.

If necessary, target eluates were concentrated using final batches of purified protein were measured by Bradford assay and typically mixed with 1x complete mini EDTA-free protease inhibitor (Roche, Basel Switzerland) and 5% glycerol prior to flash freezing in liquid nitrogen and storage in a -80°C freezer.

Cloning and Expression of Targets in Pichia and Cell Free Systems

Targets EHI_138010 and EHI_104230 were amplified from *E. histolytica* genomic DNA and cloned into pICZα and pEU-E01-LIC vectors for *Pichia pastoris* and wheat germ cell free (WGCF) expression, respectively. Later, wheat germ cell free expression of another batch of targets (EHI_066490, EHI_010740, EHI_152170, and EHI_065240) was pursued; the primers designed to facilitate ligation-independent cloning of these antigens into pEU-E01-LIC is listed in **Table 2-1**. Inserts were obtained from pUC9 vectors containing the previously codon optimized genes. The 20 μL PCR reactions consisted of 10 ng vector DNA, 0.5 μM primers, and

1x Phusion HF Mastermix. The PCR cycle used was 1) denaturation at 95°C for 30 seconds; 2) 30 amplification cycles at 95°C for 10 seconds, 59°C for 30 seconds, and 72°C for 15 seconds; and 3) final extension of 72°C for 5 minutes.

Small scale cell-free transcription and translation reactions were done as previously described by the SSGCID[187]. Once confirmed, large scale translation reactions were performed utilizing a Protemist DT II robot (Cell Free Sciences, Yokohama, Japan), and the resultant supernatant was clarified by centrifugation and purified by IMAC.

Small cultures of *Pichia pastoris* harboring targets in pICZa (Life Technologies) were grown in YPD medium overnight at 250 rpm and 30°C, and, the next day, cultures were expanded to 400 mL BGMY medium. The culture was grown for 48 hours at 30°C, with feeding of 0.5% methanol occurring 24 hours post-inoculation. The cells were then centrifuged for 10 minutes at (3500 rpm), and the media was purified via IMAC. Fractions containing visible protein were pooled and dialyzed. An aliquot was subjected to 4-12% SDS-PAGE, and three bands were excised and examined by LC-MS/MS (MS Bioworks).

Supplementary Tables

Tables 2S1 and 2S2 are provided in a separate Excel file.

TABLE 2S3. Description and Expression of pPICHOLI-2 *E. histolytica* Candidate Biomarkers.

Target Locus	Uniprot	Annotation	Whole Sequence?[a]	Residues Expressed	Molecular Weight (kDa)	Soluble?
EHI_141950	C4LW41	Putative zinc finger in N-recognin (UBR box)	No	214-424	31.0	No
EHI_136160	C4M296	Calreticulin, putative	No	189-389	30.2	No
EHI_146120	C4M3Q6	Hypothetical protein	No	16-186	25.8	No
EHI_114220	C4M429	C2 domain containing protein, putative	No	1406-1753	48.0	No
EHI_019630	C4MA47	Hypothetical protein	No	14-186	25.5	No
EHI_102270	C4MBP1	Heat shock protein 90, putative	No	1-280	38.8	No
EHI_075310	C4MBS2	Hypothetical protein	No	14-202	27.4	No
200.m00090	Q1EQ28	Clathrin heavy chain, putative	No	1-232	32.2	No
EHI_012500	Q1EQ35	Coatomer gamma subunit, putative	No	415-848	56.7	No
EHI_060860	Q76KF5	D-3-phosphoglycerate dehydrogenase, putative	No	120-299	26.7	No
EHI_110780	C4LUC5	Hypothetical protein	Yes	42-295	35.4	No
EHI_199110	C4LWV6	Lysozyme, putative	Yes	14-211	28.7	No
EHI_087690	C4LXU9	Hypothetical protein	Yes	1-388	51.3	No
EHI_038630	C4LZP7	TBC domain containing protein, putative	Yes	1-450	58.6	No
EHI_107300	C4M171	Hypothetical protein	Yes	15-296	38.6	No
EHI_192430	C4M1G3	Hypothetical protein	Yes	1-251	35.0	No
EHI_133790	C4M1M1	Hypothetical protein	Yes	16-233	31.1	No
EHI_145840	C4M3Q0	Peroxiredoxin, putative	Yes	1-233	32.7	No
EHI_047820	C4M7X3	BspA-like leucine rich repeat protein, putative	Yes	1-308	41.8	No
EHI_067080	C4MA44	Hypothetical protein	Yes	1-283	40.4	No
				Expression Success Rate (%)		0

[a]Excluding putative signal sequences and transmembrane domains

TABLE 2S4. Description and Expression of pET27b *E. histolytica* Candidate Biomarkers.

Target Locus	Uniprot	Annotation	Whole Sequence?[a]	Residues Expressed	Molecular Weight (kDa)	Soluble?
EHI_104630	**C4LZ72**	**Filamin 2**	**No**	**569-769**	**26**	**Yes[b]**
EHI_136160	C4M296	Calreticulin, putative	No	189-389	27.0	No
EHI_114220	C4M429	C2 domain containing protein, putative	No	1406-1753	41.5	No[c]
EHI_033250	**C4M6Y2**	**Polyadenylate-binding protein, putative**	**No**	**1-205**	**23.2**	**Yes**
EHI_101240	**C4M8B5**	**Nucleic acid binding related**	**No**	**1-192**	**21.5**	**Yes[b]**
EHI_102270	C4MBP1	Heat shock protein 90, putative	No	1-280	32.4	No[c]
EHI_012500	Q1EQ35	Coatomer gamma subunit, putative	No	415-848	50.4	No[c]
EHI_060860	Q76KF5	D-3-phosphoglycerate dehydrogenase, putative	No	120-299	25.2	No
				Expression Success Rate (%)		37.5

[a]Excluding putative signal sequences and transmembrane domains

[b]Target expressed at higher molecular weight than expected

[c]Target did not reach expression stage

TABLE 2S5. Description and Expression of pET28b *E. histolytica* Candidate Biomarkers.

Target Locus	Uniprot	Annotation	Whole Sequence?[b]	Residues Expressed	Molecular Weight (kDa)	Soluble?
EHI_110170	**C4LU71**	**Ubiquitinyl hydrolase, putative**	**No**	**437-1036**	**77.7**	**Yes**
EHI_010740	C4LYW1	Ubiquitin-protein ligase, putative	No	1-408	55.1	No
EHI_104630	**C4LZ72**	**Filamin 2**	**No**	**569-769**	**29.9**	**Yes[c]**
EHI_174540	C4LZK2	Tyrosyl-DNA phosphodiesterase, putative	No	1-427	57	No
EHI_064490	C4M0B8	Protein kinase, putative	No	727-977	30.0	No
EHI_136160[a]	C4M296	Calreticulin, putative	No	189-389	30.7	No
EHI_146120[a]	C4M3Q6	Hypothetical protein	No	16-186	26.3	No
EHI_114220	**C4M429**	**C2 domain containing protein, putative**	**No**	**1903-2227**	**44.9**	**Yes**
EHI_117680	C4M683	Protein kinase, putative	No	14-627	78.1	No
EHI_065240	C4M6K2	Receptor protein kinase, putative	No	1627-1916	40.7	No
EHI_075310[a]	C4MBS2	Hypothetical protein	No	14-202	27.9	No
200.m00090	**Q1EQ28**	**Clathrin heavy chain, putative**	**No**	**940-1116**	**27.7**	**Yes**
200.m00090[a]	Q1EQ28	Clathrin heavy chain, putative	No	1-232	32.7	No
EHI_152170	Q962P2	Chitinase Jessie 3, putative	No	18-202	31.9	No[c]
EHI_113200	B1N4B8	Hypothetical protein	Yes	1-726	91.9	No
EHI_110780[a]	C4LUC5	Hypothetical protein	Yes	1-445	35.9	No
EHI_092210	C4LUY6	Hypothetical protein	Yes	1-351	39.1	No[d]
EHI_126120	C4LVZ7	DNA mismatch repair protein mutL	Yes	1-615	80.1	No
EHI_096190	**C4LWD2**	**Hypothetical protein 6.t00015**	**Yes**	**1-549**	**71.9**	**Yes**
EHI_087790	C4LXV9	Hypothetical protein 20.t00054	Yes	20-398	41.4	No[d]
EHI_118130	**C4LYZ7**	**C2 domain protein, putative**	**Yes**	**1-389**	**49.5**	**Yes[c]**
EHI_135140	C4M100	Methyltransferase TRM13, putative	Yes	1-468	61.7	No
EHI_145840	**C4M3Q0**	**Peroxiredoxin, putative**	**Yes**	**1-233**	**33.2**	**Yes**
EHI_028930	C4M4V7	Cyst wall-specific glycoprotein Jacob	Yes	22-151	22.4	No
EHI_128140	C4M532	Hypothetical protein 45.t00028	Yes	1-659	84.8	No
EHI_122810	C4M5M7	Putative zinc finger in N-recognin (UBR box), putative	Yes	1-690	88.6	No
EHI_033250	**C4M6Y2**	**Polyadenylate-binding protein, putative**	**Yes**	**1-521**	**66.1**	**Yes**
EHI_166420	C4M7E8	Cyclin, putative	Yes	1-400	54.2	No
EHI_158620	**C4M7G6**	**Opioid growth factor receptor (OGFr) conserved region**	**Yes**	**1-402**	**54.9**	**Yes[c]**
				Expression Success Rate (%)		31.0

[a]Genes excised from pPICHOLI-2 vector and re-cloned into pET28b

[b]Excluding putative signal sequences and transmembrane domains

[c]Target expressed at a larger molecular weight than expected

[d]Target did not reach expression screen stage

TABLE 2S6. Description and Expression of *E. histolytica* Candidate Biomarkers in SSGCID Tier One.

Target Locus	Uniprot	Annotation	Whole Sequence?[a]	Residues Expressed	Molecular Weight (kDa)	Soluble?
EHI_135140	C4M100	Methyltransferase TRM13, putative	--[b]	--	--	No
EHI_107300	C4M171	Hypothetical protein	--[b]	--	--	No
EHI_146120	C4M3Q6	Hypothetical protein	--[b]	--	--	No
EHI_065240	C4M6K2	Receptor protein kinase, putative	--[b]	--	--	No
EHI_158620	C4M7G6	Opioid growth factor receptor (OGFr) conserved region	--[b]	--	--	No
EHI_110170	C4LU71	Ubiquitinyl hydrolase, putative	No	1-739	87.1	No
			No	740-937	24.4	No
			No	938-1143	25.2	No
EHI_141950	C4LW41	Putative zinc finger in N-recognin (UBR box)	No	791-926	17.0	No
EHI_010740	C4LYW1	Ubiquitin-protein ligase, putative	No	515-1020	60.8	No
			No	1021-1499	58.5	No
			No	1500-2126	75.3	No
EHI_174540	C4LZK2	Tyrosyl-DNA phosphodiesterase, putative	No	1-413	49	No
			No	414-1323	107	No
EHI_064490	C4M0B8	Protein kinase, putative	No	Unknown	--	No
EHI_114220	C4M429	C2 domain containing protein, putative	No	913-1610	83.1	No
			No	1400-1801	49.3	No
			No	1611-1935	38.9	No
			No	1900-2332	51.5	No
EHI_117680	C4M683	Protein kinase, putative	No	1-201	24.7	No
200.m00090	**Q1EQ28**	**Clathrin heavy chain, putative**	**No**	**1-319**	**36.0**	**Yes**
			No	**400-736**	**39.0**	**Yes**
			No	753-886	16.0	No
			No	848-1013	20.0	No
			No	1010-1158	18.0	No
			No	**1114-1448**	**39.0**	**Yes**
EHI_012500	Q1EQ35	Coatomer gamma subunit, putative	No	298-848	64.5	No
EHI_152170	Q962P2	Chitinase Jessie 3, putative	No	24-336	35.6	No
EHI_110780	C4LUC5	Hypothetical protein	Yes	1-445	50.8	No
EHI_092210	C4LUY6	Hypothetical protein	Yes	1-351	40.1	No
EHI_096190	C4LWD2	Hypothetical protein 6.t00015	Yes	1-549	71.9	No
EHI_199110	C4LWV6	Lysozyme, putative	Yes	17-211	24	No
EHI_087690	C4LXU9	Hypothetical protein	Yes	1-388	45.8	No
EHI_087790	C4LXV9	Hypothetical protein 20.t00054	Yes	20-398	41.4	No
EHI_118130	C4LYZ7	C2 domain protein, putative	Yes	1-389	42	No
EHI_104630	**C4LZ72**	**Filamin 2**	**Yes**	**1-857**	**96.4**	**Yes**
EHI_038630	C4LZP7	TBC domain containing protein, putative	Yes	1-450	58.6	No
EHI_006810	**C4M0F4**	**14-3-3 protein 3**	**Yes**	**1-240**	**27.5**	**Yes**
EHI_015290	**C4M0W7**	**C2 domain protein, putative**	**Yes**	**1-281**	**30**	**Yes**
EHI_192430	**C4M1G3**	**Hypothetical protein**	**Yes**	**1-251**	**29.6**	**Yes**
EHI_133790	C4M1M1	Hypothetical protein	Yes	24-213	23	No
EHI_136160	**C4M296**	**Calreticulin, putative**	**Yes**	**14-389**	**45.1**	**Yes**
EHI_109890	**C4M3K8**	**Chitinase, putative**	**Yes**	**16-577**	**58.0**	**Yes**
EHI_028930	C4M4V7	Cyst wall-specific glycoprotein Jacob	Yes	22-151	22.4	No
EHI_128140	C4M532	Hypothetical protein 45.t00028	Yes	1-659	84.8	No
EHI_122810	C4M5M7	Putative zinc finger in N-recognin (UBR box), putative	Yes	1-690	88.6	No
EHI_033250	**C4M6Y2**	**Polyadenylate-binding protein, putative**	**Yes**	**1-521**	**59.8**	**Yes**
EHI_166420	**C4M7E8**	**Cyclin, putative**	**Yes**	**1-400**	**47.9**	**Yes**
EHI_101240	**C4M8B5**	**Nucleic acid binding related**	**Yes**	**1-843**	**95.6**	**Yes**
EHI_067080	C4MA44	Hypothetical protein	Yes	1-283	36.0	No
EHI_019630	C4MA47	Hypothetical protein	Yes	14-186	25.5	No
				Expression Success Rate (%)		26.3

[a]Excluding putative signal sequences and transmembrane domains

[b]Target did not reach expression screen

CHAPTER III: SELECTION OF YEAST-DISPLAYED SCFV AND GENERATION OF NANOYEAST-SCFV AGAINST *ENTAMOEBA HISTOLYTICA* ANTIGENS

Introduction

Heterologous peptide display by *Saccharomyces cerevisiae* has been regarded as a very useful tool in protein engineering and evolution since its inception nearly 20 years ago. This technology has been employed to display a number of proteins, including green fluorescent protein, human interleukin-2, and human epidermal growth factor, as fusions to the *S. cerevisiae* a-agglutinin mating receptor subunit Aga2p[202]. The widest use has been the display of recombinant antibodies, including scFv, Fabs, and a combination of the two called single-chain Fabs[82,203,204]. Yeast-displayed antibody combinatorial libraries have been constructed from both non-immune and immune sources, and they have produced affinity reagents specific to a wide number of disease targets[82,84,203,205].

Yeast antibody display selection methodology is rapid and simple. Expression and subsequent display of Aga2p-heterologous protein fusions in this system is inducible by a galactose promoter. With their size and thousands (10^4-10^5) of displayed antibody copies, yeast in combinatorial libraries can be labeled with antigen in solution, overcoming avidity effects and epitope biases seen with antigen panning during phage display selections[82]. However, displaying yeast can be panned on solid surfaces, as necessary[206]. Selection begins with enrichment by magnetics beads, which brings the library diversity down to a sortable level[207,208]. The resultant outputs can then be further enriched by fluorescently activated cell sorting (FACS), which can separate the yeast by desired properties, including size, antibody expression, and antigen

binding[85,209]. With FACS, the operator has excellent control of which antibody populations are collected.

. Yeast-displayed scFv (yeast-scFv) have been likened to ready-made beads with immunoprecipitation capabilities[8686]. Without further manipulation, yeast-scFv have successfully enriched antigen via immunoprecipitation for Western blot or mass spectrometry, captured antigen in an immunofluorescence microscopy assay, and mapped epitopes on their target antigens[197,206,208,210]. It has been long known that the biophysical properties of scFv, such as affinity and stability, can be measured while they are still attached to the yeast cell wall[85,211].

In spite of those beneficial properties, there are a plethora of diagnostic assays and biomedical applications that require soluble antibody and for which whole yeast-scFv cells would be too large. Unfortunately, endeavors in converting displayed, recombinant antibodies into soluble molecules have large failure rates and create a significant bottleneck in the widespread utilization of yeast antibody display, as well as phage display. Secreting recombinant antibodies post-selection and characterization adds another laborious step to antibody development that often results in low, functional yield[102,212]. Prokaryotic expression of these mammalian-derived antibodies can result in misfolded antibodies and loss of activity[81,212]. Despite that yeast systems are more comparable with mammalian expression, processing and folding bottlenecks can occur in the ER and Golgi, and, hence, expression yields can vary widely[100]. Additionally, the Aga1-Aga2p linkage to yeast cell walls may exert constraints on scFv structure that are necessary for proper recognition and binding. When the scFv is cloned and expressed separately from that linkage, it is likely to fold differently, resulting in altered (usually inferior) function. Essentially, when the displayed scFv is removed from the

environment in which it was selected to function, it rarely functions properly. New methods that reproduce and maintain the selected environment must be considered.

Herein, I describe the yeast-scFv selection of 34 clones against seven *Entamoeba histolytica* antigens, which were described in Chapter 2, and the application of these reagents while maintained in their selection environment. Our investigation is the first to utilize yeast-scFv display for selection against *Entamoeba* antigens and to utilize multiple, rationally chosen *Entamoeba* biomarkers as selection targets. We also report continued investigation on the lyophilization and subsequent dry storage stability of whole yeast-scFv, an endeavor that has added convenience to the affinity reagent development process[197]. More importantly, stable storage allowed for us to develop a novel class of reagent, called nanoyeast-scFv, which are scFv attached to fragments of the yeast cell wall on which they were originally displayed. We demonstrate these reagents function in emerging diagnostic technologies, including label-free electrochemical biosensors and surface enhanced Raman spectroscopy (SERS), as well as in the classical ELISA format with superior sensitivity and sensitivity. The nanoyeast-scFv impedance and SERS was led by the laboratory of Matt Trau, University of Queensland, and the lyophilization stability study was headed by the diagnostic program of PATH, both with input and support from this candidate.

Results

Yeast-scFv Selections

Yeast selections were conducted at least once on 15 out of the 23 soluble antigens described in Chapter 2 (**Table 3-1**). The process was adapted from previous descriptions; the

TABLE 3-2. The Success of Yeast-scFv Selections Against *Entamoeba histolytica* Candidate Biomarkers

Antigen	Selection Success?	Unique Clones
EHI_115350	Yes[a]	1
EHI_182030	Yes	3
EHI_006810	Yes	3
EHI_033250	Yes	7
EHI_104630	Yes	3
EHI_044500	Yes	3
EHI_070730	Yes	14
EHI_142000	No[a]	--
EHI_000780	No[a]	--
EHI_015290	No	--
EHI_158620	No	--
200.m00090	No	--
EHI_109890	No	--
EHI_101240	No	--
EHI_138010	No	--

[a]Reported in source [197]

largest change was culturing the second round output and checking for antigen binding before proceeding to third round cell sorting [82,208].

Clones binding to non-biotinylated antigens were obtained for a total of seven antigens, with most selections undergoing only a single FACS enrichment. Selections against antigen EHI_115350, as well as antigens EHI_142000, and EHI_000780, was previously reported in great detail[197]. There were numerous reasons for selection failure. The most frequently observed was that entire selection outputs were comprised of yeast-scFv binding non-specific or biotin-dependent epitopes. Two antigens, EHI_158620 and EHI_015290, did not have antigen-binding populations to sort during the third selection round, possibly due to a lack of immunogenicity. Additionally, a small subset of antigens were selected against using epitope tag antibodies in lieu of biotin labels during the process, but these experiments entirely failed, most often due to selection of anti-antibody scFv (data not shown). The antigens utilized in those experiments, as

well as those only selected against in single attempts, may be worth pursuing in future selections to truly ensure no antigen-specific yeast-scFv were missed.

Amongst successful selections, more than one unique scFv were often isolated, as determined by BstNI digest of scFv DNA (**Table 3-1**). **Figure 3S1** shows the diverse BstNI digest patterns of EHI_033250-selected clones, as an example. The specificity to a non-biotinylated epitope was often confirmed by flow cytometry stains with epitope tag or antigen-specific antibodies as in **Figure 3S2**, where yeast-scFv clones specific to FLAG-tagged EHI_104630 are stained with antigen and anti-FLAG primary antibody. Additionally, as biotinylated antigen was scarce after the third selection round, yeast-scFv binding EHI_044500 were identified by pairing with an antigen-specific chicken polyclonal antiserum (data not shown).

Yeast-scFv Lyophilization and Stability

A previous stability study from the AMPP consortium demonstrated that yeast-displaying scFv could be lyophilized and stored four to six weeks dry with no loss of scFv activity upon reconstitution[197]. We undertook a new trial to test additional conditions that could further extend yeast-scFv shelf-life and permit storage in relevant temperatures. Confirmed antigen-binding yeast-scFv clones were lyophilized whole in a solution of known stabilizers. The dried cakes were stored in sealed vials and at three temperatures (25°C, 37°C, and 45°C) for up to a year. Yeast-scFv clones 350-E2, 030 C, and 14-3-3 D, each specific to a different antigen, were reconstituted, stained and analyzed by flow cytometry monthly to monitor scFv binding activity[213].

TABLE 3-2. Antigen Binding Activity of Lyophilized Yeast-scFv Clone 350-E2 Over One Year.

		350-E2					
		25°C		37°C		45°C	
		350 Ag[a]	780 Ag	350 Ag	780 Ag	350 Ag	780 Ag
Days Post-Lyophilization	14	+	-	+	-	+	-
	21	+	-	+	-	+	-
	60	+	-	+	-	+	-
	180	+	-	+	-	+	-
	360	+	-	+	-	+	-

[a] "Ag"= antigen; antigens 350 and 780 are the cognate and non-cognate antigens, respectively.

Table 3-2 shows yeast-scFv 350-E2's general results across days 14-360 post-lyophilization at the three tested storage temperatures[213]. Clones 030 C and 14-3-3 D had similar results. All three clones had antigen-binding activity over a year's storage at all temperatures, up to 45°C. This binding remained specific, as the clones did not bind non-cognate antigens at any time over the study.

Nanoyeast-scFv Electrode Assay

Figure 3-1 is a generalized schematic of the nanoyeast-scFv assays created thus far. In all cases, the nanoyeast-scFv are selectively enriched and tethered at a solid surface via an anti-hemagglutinin (HA) monoclonal antibody, which is specific to the HA epitope tag that has been engineered in the scFv of this library[82]. The attachment of the anti-HA antibody to the surface varies: biotinylated BSA and streptavidin capture biotinylated anti-HA in the electrode assays, biotinylated polyethylene glycol thiol (Bio-PEG-SH) and streptavidin performed the same duty in the SERS assay, and, in the ELISA, the anti-HA is directly coated to the plate's plastic surface. Then, the yeast-scFv, having been fragmented from whole cells into nanoyeast-scFv, is placed in the assay and is bound by immobilized anti-HA. Once in place, the scFv capture

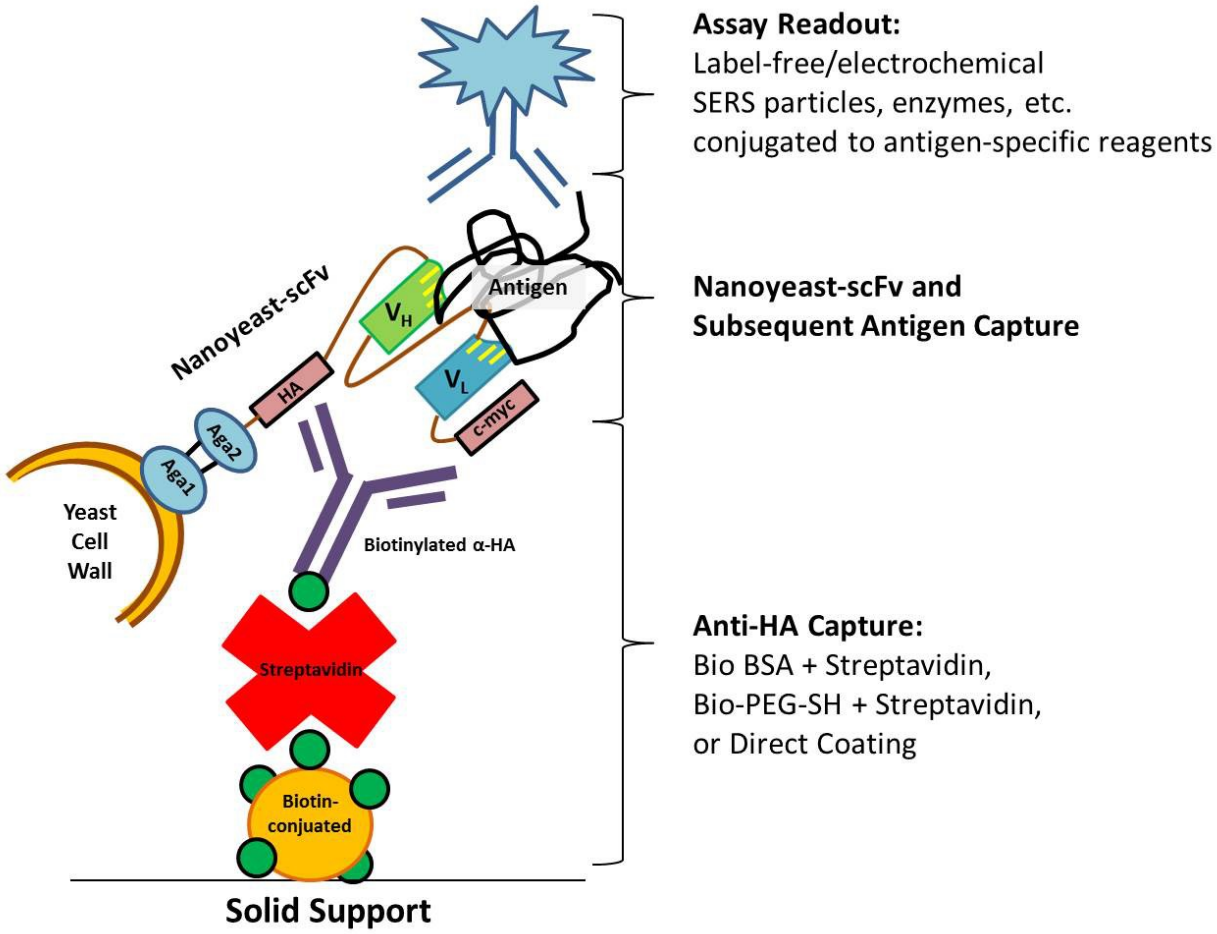

Figure 3-1. A schematic of the nanoyeast-scFv assays described herein. These assays can be generalized into three parts: 1) tethering of anti-hemagglutinin antibody to a solid support; 2) capture of nanoyeast-scFv fragments via a hemagglutinin tag that was engineered into the scFv during library construction[82], followed by capture of cognate antigen by the nanoyeast-scFv itself; and 3) readout of antigen capture, either by label-free electrochemical signals or by reporter-conjugated antibodies specific to antigen.

antigen as they would displayed on whole yeast cells. From there, antigen capture can be quantified by antibody-conjugated fluorescent or enzymatic readout molecules or by label-free electrochemical signals.

The first assays with nanoyeast-scFv utilized the surfaces of gold electrodes to immobilize reagents in a step-wise fashion, with a final incubation in antigen solution. Once electrode preparation was complete, electrical impedance spectroscopy and dual-pulse voltammetry were conducted with electrodes dipped in a redox solution[214–216]. The impedance of

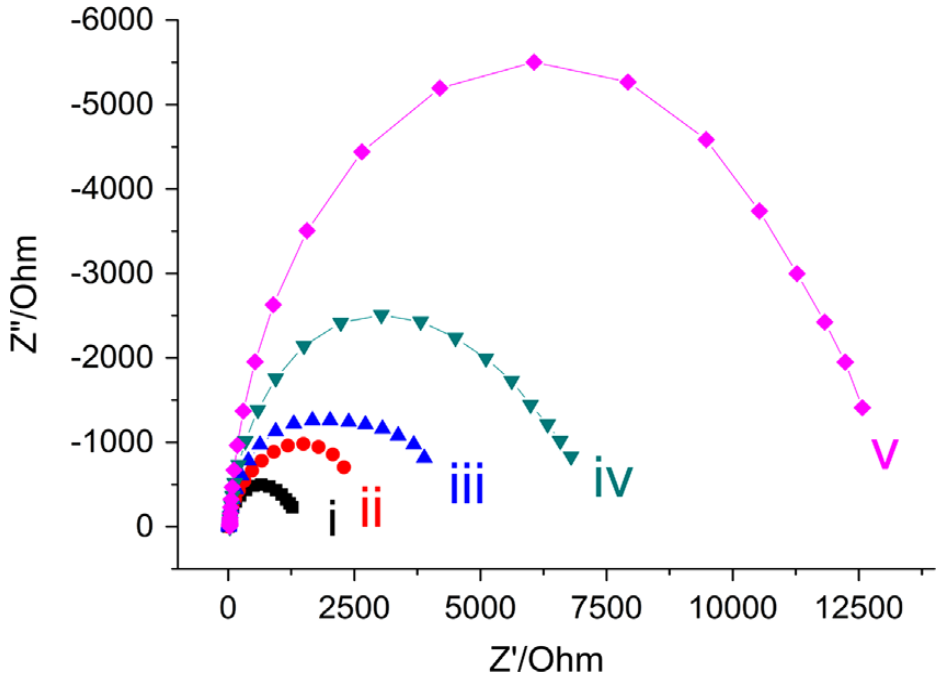

Figure 3-2. Nyquist plot demonstrating the sequential build-up of assay reagents on the surface of gold screen-printed electrodes. The area under each curve represents the average resistance measured on the surfaces of three independent electrodes when dipped in an electrolyte solution (2.5 mM $K_3[CN_6]$, 2.5 $K_2[CN_6]$, and 0.1 M KCl in PBS). The curves represent, as follows: **(i)** biotinylated BSA alone; **(ii)** biotinylated BSA and streptavidin; **(iii)** biotinylated BSA, streptavidin, and biotinylated anti-hemagglutinin antibody; **(iv)** biotinylated BSA, streptavidin, biotinylated anti-hemagglutinin antibody, and nanoyeast-scFv prepared from clone 350-E2; and **(v)** biotinylated BSA, streptavidin, biotinylated anti-hemagglutinin antibody, 350-E2 nanoyeast-scFv, and 350 antigen at 1000 pg/mL concentration. Reprinted from [216]. Copyright 2014 Elsevier. doi:10.1016/j.bios.2013.12.043

electron transfer on the surface of screen-printed gold electrodes increased with each addition of reagent, and the greatest increase occurs with the addition of cognate antigen 350 (**Figure 3-2**)[216]. Such an increase did not occur with the addition of non-cognate antigen, and this format was able to detect cognate antigen in buffer down to 10 pg/mL[216].

The most significant finding from these experiments was the sensitivity and specificity of nanoyeast-scFv-modified electrodes as revealed by analysis of a complex stool matrix[216]. A complex biological matrix derived from *E. histolytica*-negative stool and diluted 1:5 in phosphate buffered saline (PBS) was spiked with antigens 350 and 030, and electrodes with cognate nanoyeast-scFv (derived from clones "350-E2" and "030 L") were dipped and mixed in the assay's incubation step, as before. There was a specific increase in the electrodes' resistance

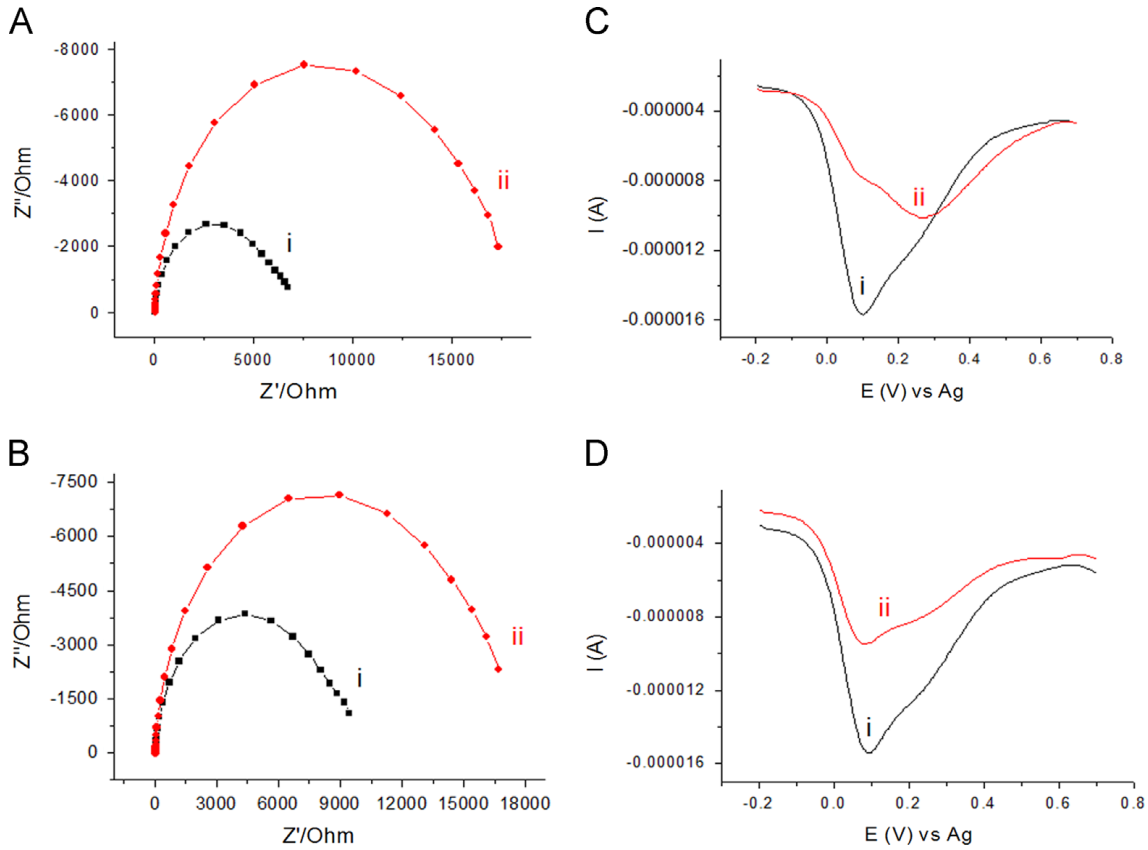

Figure 3-3. Nyquist plots and DPV measurements of nanoyeast-scFv 350-E2 (A & C) and 030 L (B & D) tethered to gold screen-printed electrodes and assayed in a complex stool matrix. The black lines and curves (i) represent 1:5 blank stool matrix with no antigen, whereas the red lines and curves (ii) represent 1:5 stool matrix spiked with 500 pg/mL 350 antigen (A & C) or 030 antigen (B & D). Reprinted from [216]. Copyright 2014 Elsevier. doi:10.1016/j.bios.2013.12.043

in the presence of the cognate antigens at 500 pg/mL concentration, when compared to similarly prepared electrodes dipped in blank stool matrix (**Figure 3-3 A&B**)[216]. Nanoyeast-scFv prepared from 030 L showed more non-specific binding than 350-E2-derived nanoyeast-scFv. Additionally, dual-pulse voltammetry measurements mirrored these results, showing an ablation of redox-reaction when the electrode had incubated in stool containing cognate antigen, versus incubation in a blank stool control (**Figure 3-3 C&D**).

Nanoyeast-scFv SERS Assay

The same nanoyeast-scFv reagents tested with the screen-printed electrodes were employed in a SERS duplex assay[217]. As before, they were captured by biotinylated anti-HA

antibody, which was itself tethered to the channels of a microfluidic device by streptavidin and Bio-PEG-SH molecules. The readout of this assay was facilitated by two reporter molecules, MBA and MBNA, which generate unique spectra under a single excitation wavelength, allowing detection of the individual antigens within the same mixture. These molecules were coated onto gold nano-particles and encapsulated in silica, and each reporter was conjugated to a polyclonal antibody recognizing one of the nanoyeast-scFv's cognate antigens. The MBNA-containing SERS particles were attached to anti-350 rabbit polyclonal antibody, while MBA-containing SERS particles were conjugated to a polyclonal chicken antibody with specificity to antigen 030.

Each nanoyeast-scFv showed sensitive detection of cognate antigen in this assay format: 350 antigen was detected down to 1 pg/mL in buffer, whereas 030 was detected down to 10 pg/mL (data not shown)[217]. The reagents were then tested in a 3-channel format, with one channel each containing one of the nanoyeast-scFv alone and a third channel containing both at a 1:1 concentration. All three channels were incubated with both antigens (500 pg/mL 350 and 5 ng/mL 030) and antibody-conjugated SERS particles. Specific antigen detection was achieved with little cross-reaction (**Figure 3-4**). The unique spectra of the reporter molecules is shown in the channels harboring their matching scFv, whereas the third channel containing both nanoyeast-scFv displays a combined spectra of both reporters' signatures. The higher cross-reaction seen of the 350-E2 nanoyeast-scFv with the 030 chicken antibody reporter was most likely due to remnant non-specific binding of the chicken antibody to yeast cell wall components, although this was not tested directly.

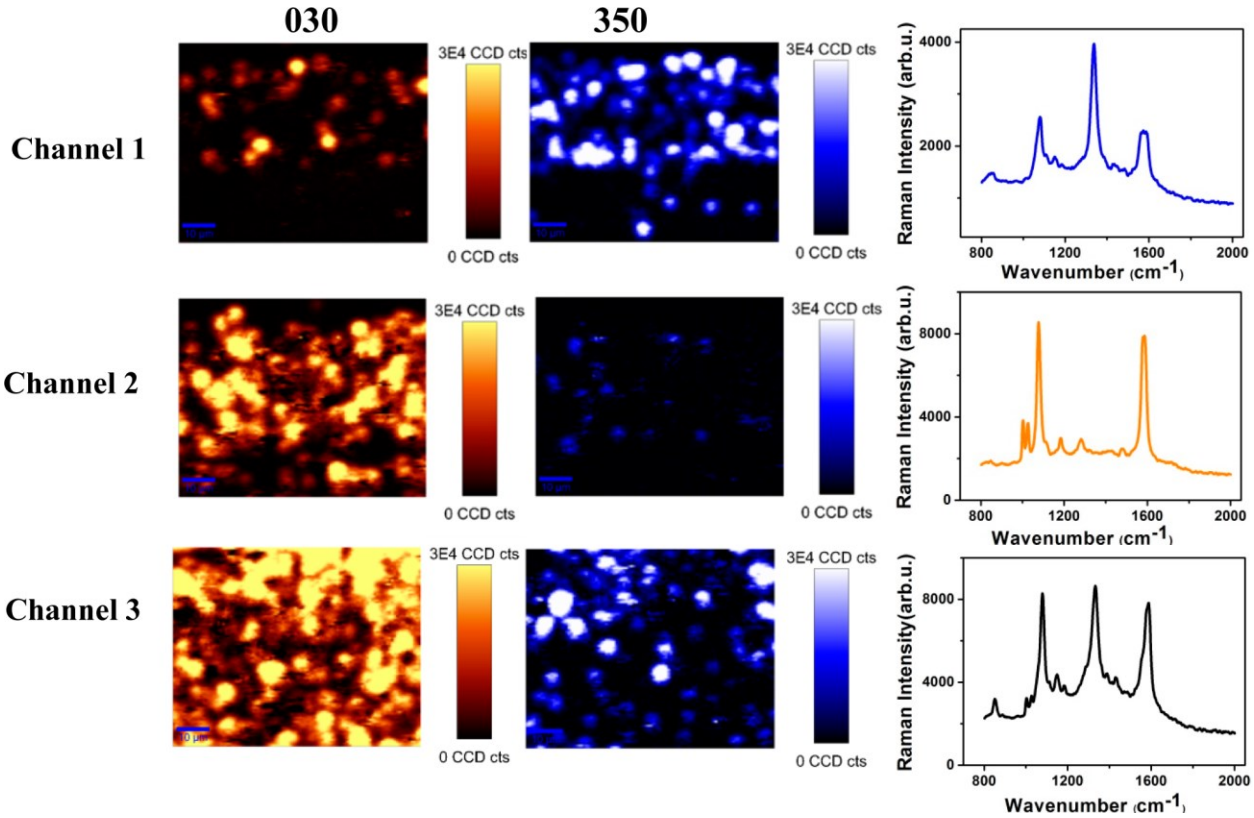

Figure 3-4. SERS false-color microscopic images and spectra obtained from a 3-channel nanoyeast-scFv antigen capture assay. The intensity scales for the false-color images are displayed to the right of every image. Channel 1 of this device contained nanoyeast-scFv derived from clone 350-E2, channel 2 contained nanoyeast-scFv from clone 030 L, and channel 3 contained both in a 1:1 ratio. All three channels received both antigens, 350 and 030, as well as the distinct SERS reporter particles conjugated to their polyclonal antibodies. Reprinted with permission from [217]. Copyright 2014 American Chemical Society. http://pubs.acs.org/doi/full/10.1021/ac5027012

Nanoyeast-scFv ELISA

With the success of the above nanoyeast-scFv assays, we investigated adaption of these reagents to an ELISA format. It was observed that, like the electrode and SERS assays, nanoyeast-scFv could be captured in anti-HA antibody-coated microplate wells. This binding was confirmed by detection of the scFv's other tag, c-myc (**Figure 3-1**), with primary antibody and HRP-conjugated secondary, and it was dependent on the concentrations of anti-HA antibody and nanoyeast-scFv placed in the wells (**Figure 3S3**). No binding was seen in the absence of the anti-HA antibody. We also tested nanoyeast-scFv prepared from yeast-scFv cells that were not cultured in galactose and thus not induced to display scFv. There was some non-specific binding

with higher anti-HA concentration and the non-induced nanoyeast-scFv control (highest signal at OD450 a little more than 0.5), which suggests that there is leaky expression of scFv on in this yeast clone in the non-induced condition.

Next, a single yeast-scFv clone, Jacob D9, was employed as nanoyeast-scFv to detect its cognate antigen in this format. Jacob antigen was titrated across prepared wells from 50 nM down to 39 pM (**Figure 3-5**). When probed with anti-c-myc antibody, wells containing induced reagent had high absorbance, compared to similarly-probed wells with uninduced reagent. Antigen detection with this nanoyeast-scFv and with primary antibody (1D3 shown below) was specific, as there was no signal with the *E. dispar* version of the antigen (EdJacob) at high concentration. Compared to antigen titration with uninduced yeast, the limit of detection was 1.6

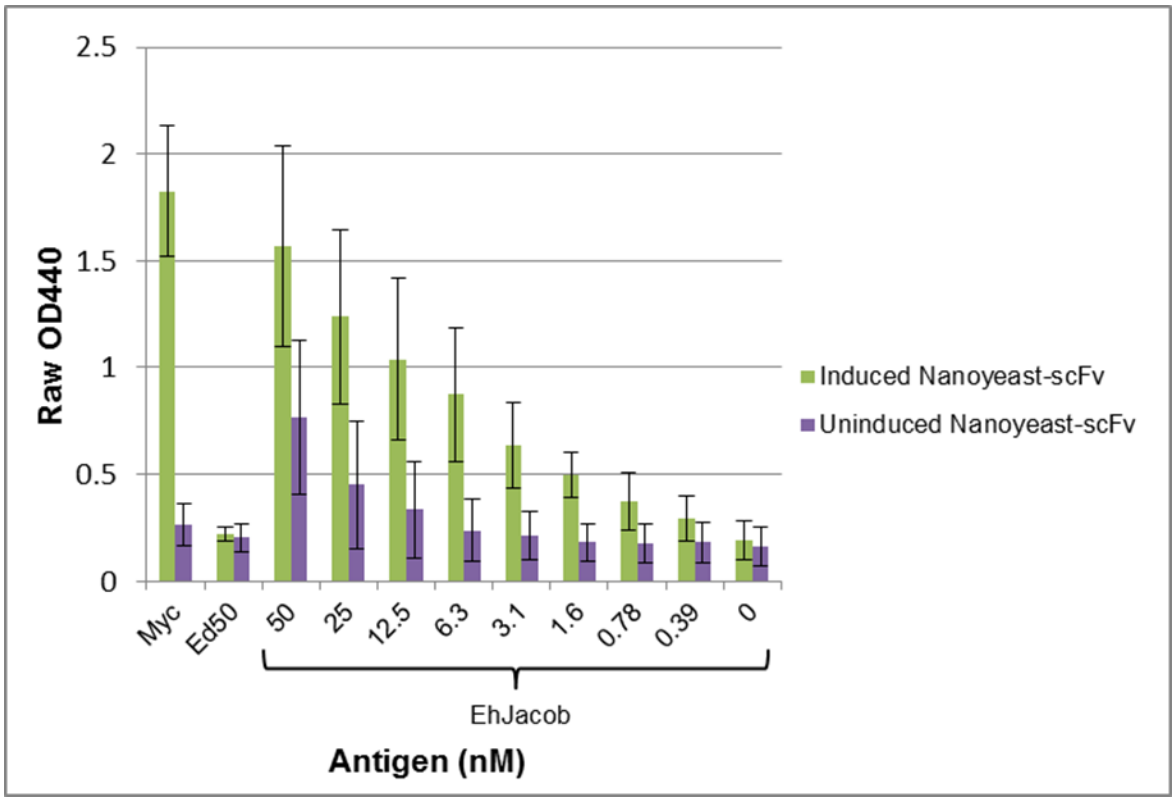

Figure 3-5 Antigen capture by nanoyeast-scFv in an ELISA format. Induced and uninduced Jacob D9 yeast-scFv were generated into nanoyeast-scFv at 0.5 cake/mL concentration and were captured in anti-HA antibody-coated microplate wells. Capture of the nanoyeast-scFv was determined by probing with anti-c-myc primary antibody ("Myc"). The *Entamoeba dispar* Jacob antigen at 50 nM ("Ed50") was used as a non-cognate antigen control. Cognate *Entamoeba histolytica* antigen was assayed from 50 nM down to 39 pM. These results represent the average and standard deviations of three independent

nM. However, there was concerning non-specific binding in the wells at the start of the titration. At this time, it is unknown whether there is a small level of scFv displayed on the non-induced yeast that is causing this cross-reaction or if there is interaction amongst the disrupted yeast cell wall, antigen, and other assay components. Additionally, between three assay runs, the signals vary considerably, which could suggest that preparation of the nanoyeast-scFv is yet to be optimized and standardized between operators and time of assays.

Discussion

In this body of work, we have successfully selected whole-cell, yeast-scFv affinity reagents against *E. histolytica* candidate biomarkers, created a protocol that stabilizes these reagents for up to a year, and generated and validated a novel class of scFv affinity reagents, which have been named nanoyeast-scFv. Selections were conducted routinely, with no significant alterations to previously developed protocols[82,208]. I achieved a good success rate (7/15, 47.7%) in obtaining antigen-binding clones against our targets, although some targets produced only non-specific or biotin-binding clones or even no clones at all. The success rate could improve with repeated selections against some antigens, possibly by selecting a wider gate during FACS or by screening a larger proportion of sorted clones.

We improved a previous lyophilizaton protocol of whole yeast-scFv, extending the dry shelf-life 8-13-fold[197]. Some factors that likely contributed to this success are the incorporation of known stabilizing compounds in the lyophilization media, as well as higher-quality equipment, vials, and sealing procedures. Once induced in a galactose-rich culture, yeast-scFv clones only maintain display for two to three weeks, and so this lyophilization protocol overcomes limitations encountered in yeast-scFv characterization and storage.

Furthermore, this improved protocol has allowed us to generate a novel class of affinity reagent. Rather than recloning a selected scFv's gene into a bacterial or yeast secretion vector to obtain soluble reagent, which usually results in loss of activity, we have instead fragmented lyophilized yeast-scFv to generate cell wall fragments with displayed scFv intact. The rationale is that secreting scFv removes them from the environment from which they were selected for function, fused to the yeast cell wall protein. The fusions may exert constraints that are required for proper conformation and antigen binding. As demonstrated above, nanoyeast-scFv consistently remain active even after dry storage and harsh fragmentation, consistent with the idea that some semblance of the selection environment is beneficial. Additionally, generation of nanoyeast-scFv is much simpler, more rapid, and less costly than re-cloning scFv into a new vector. To date, we have demonstrated this concept with three different yeast-scFv clones. More work will determine the consistency with which yeast-scFv clones will produce stable, function nanoyeast-scFv.

An exciting aspect of this work is demonstration of nanoyeast-scFv functionality in a variety of novel technologies, in addition to a traditional ELISA-like format. Purified scFv have already been utilized in both electrochemical biosensing and SERS[214,218]. Both assay formats were also successful when nanoyeast-scFv were incorporated, with specific detection of cognate and analytical specificities typically as low as 1-10 pg/mL[215–217]. With use of distinct SERS particles conjugated to secondary reagents, the SERS assay conferred the additional benefit of duplex antigen detection in that sensitivity range[217]. Additionally, in the electrochemical format, two antigens were specifically detected by their cognate nanoyeast-scFv in a complex stool matrix at 500 pg/mL, which is half the limit of detection of commercial *E. histolytica* Gal/GalNAc stool ELISAs[150]. Furthermore, after these studies, another investigation showed that

whole yeast-scFv could be tethered directly to gold electrodes via gold binding protein (GBP) and 6xHis tags and detect their cognate antigens as in our assays[219]. Though that assay was four to five magnitudes less sensitive (and it was conducted with different scFv and antigens), it validated our efforts in utilizing whole yeast-scFv and their derivatives in assays[216,217,219].

This is not the first time alternative affinity reagents have been selected for *E. histolytica* antigens. Fabs have been generated against the Gal/GalNAc lectin, some with diagnostic or therapeutic potential[220–223]. Additionally, an scFv against EHI_136160, one of our own candidate biomarkers, was selected against recombinant antigen via phage display and was employed in staining *E. histolytica* trophozoites[192]. With multiple scFv isolated per target, it is hoped that at least one reagent will demonstrate similar activity against native antigen. As we continue developing and characterizing nanoyeast-scFv, this will certainly be investigated.

Our initial work on nanoyeast-scFv represents progress toward bypassing the limitations seen with yeast antibody display. Future directions with these reagents include improving and standardizing their generation (the necessity of which is suggested by the ELISA assay) and developing methods to enrich these reagents for use in soluble applications. The overall goal is that nanoyeast-scFv will arise as a fully functional, easy-to-produce, and affordable class of affinity reagents with broad applicability in biomedical research and diagnostic application.

Materials and Methods

Biotinylation of Antigens

Antigens for this study were typically biotinylated with 3-20 molar excess of EZ Link Sulfo-NHS-LC-Biotin or NHS-PEG4-Biotin (Life Technologies, Carlsbad, CA). A subset of antigens contained the BCCP biotinylation site and was biotinylated with use of a BirA-500 kit,

per manufacturer's instructions (Avidity LLC, Aurora, CO). Excess biotin was removed by Zeba columns (Life Technologies) or dialysis. Biotinylation was confirmed by Western blots probed by streptavidin-HRP reagent (Sigma Aldrich, St. Louis, MO).

Generation of Polyclonal Antibodies

Polyclonal antisera to recombinant *E. histolytica* antigens EHI_115350 (350) and EHI_044500 (Jacob) were each generated by Cocalico Biologicals, Inc. (Reamstown, VA)[186,197]. A chicken polyclonal IgY preparation specific to recombinant antigen EHI_182030 (030), as well as antigens EHI_101240, EHI_006810, and EHI_070730, was produced by Aves Labs (Tigard, OR). Chicken polyclonal IgY antibody was also generated against a mixture of Jacob, 350, EHI_142000, EHI_109890, EHI_000780, and the LecA antigen (Techlab, Blacksburg Virginia)[224]. Murine monoclonal antibodies to EHI_044500 were produced by the Antibody Technology Core at the Fred Hutchinson Cancer Research Center (Seattle, WA).

Generation of Stool Matrix

Human stool samples were collected with informed consent by the International Center for Diarrhoeal Disease Research, Bangladesh (ICDDR,B). Approval for this collection was obtained from the ICDDR,B and the University of Washington. After same-day transport to the ICDDR,B and diagnostic inspection, the stool samples were stored and shipped at -80°C until further use.

Five *E. histolytica*-negative stool samples were thawed and combined at approximately a 1:1 ratio with 1x PBS. The mixture was aliquoted into 2 mL screwcap tubes and centrifuged at 7300 x g for 5 minutes. The supernatant was drawn off, realiquoted into screwcap tubes, and

placed in a 95°C water bath for 20 minutes. Once heated, the tubes were placed on ice. Disinfection was confirmed when a streak from every tube produced no growth on a nutrient agar and a tryptic soy agar plate after ≥72 hours at 37°C.

Selection of Yeast-scFv

A non-immune scFv library was provided by K. Dane Wittrup[82]. Selection protocols were generally adapted from previously published methods, and the particular protocol for selections of EHI_115350, EHI_000780, and EHI_142000 are previously described[82,197,208].

The first two rounds of selection, which enrich yeast-scFv via a magnetic column, were conducted as previously described[82,197,208]. Unlike those published methods, enrichment by FACS was delayed until the magnetically-enriched cells could be further expanded in culture. At the end of the second selection round, yeast eluted from the magnetic column were grown overnight in SD-CAA to at least four times their original number. The next day, yeast representing four times the amount of previous output were harvested, washed twice, and divided into 4 aliquots. For 45 minutes, two of these aliquots were incubated with 500 µL blank yeast wash buffer (YWB)(1x PBS + 0.05% bovine serum albumin) , and the other two were mixed with 500 µL of 100 nM biotinylated antigen. Next, all aliquots were washed twice and resuspended in 500 µL of 1.25 µg/mL phycoerythrin-conjugated streptavidin (SA-PE, Life Technologies) and 5 µg/mL FITC-conjugated anti-c-myc antibody (ICL Laboratories, Portland, OR) for 30 minutes. Finally, they were washed three times, and one aliquot each from the two subsets (with antigen, without antigen) was analyzed by a two-color flow cytometry protocol on a Coulter Epics XL MCL (Beckman Coulter, Inc., Brea, CA).

Depending on the flow cytometry results, the remaining aliquots of selected yeast were strained through the strainer snap cap of Falcon round bottom tubes (Corning, Corning, NY) and subsequently sorted on a Beckman Aria sorter (Beckman Coulter). Collections ranged from 5000-50000 cells. With some selections (involving antigens EHI_115350 and EHI_182030), sorted cells were grown, restained with lower concentrations of antigen (1-10 nM), and sorted a second time[197]. After final sorts, the output was dilution plated on selective growth agar and allowed to grow at 30°C for 72 hours.

Yeast-scFv Specificity Assays

Colonies arising from sorted yeast were picked and placed in deep wells containing 1 mL SD + CAA and grown overnight shaking at 30°C. The next day, cultures were inoculated 1:10 into 900 µL induction media and grown overnight at 20°C.

Once induced, fifty microliters of yeast-scFv (regardless of density) was plated into two wells of a Falcon V-bottom microplate (Corning). Wells were washed by resuspending yeast in 200 µL YWB, centrifuging the plate at 4300 rpm for 1 minute, and discarding the buffer. With 2-3 washes in between each step, the yeast was incubated with 100 nM biotinylated cognate or non-cognate antigen for 45 minutes at room temperature, followed by 1.25 µg/mL SA-PE and 5 µg/mL α-c-myc-FITC for 30 minutes.

To identify clones able to bind non-biotinylated antigen, screening was conducted similarly, except that SA-PE was replaced with anti-epitope tag or anti-antigen primary antibody and appropriate secondary antibody stains for 45 minutes each. Murine anti-FLAG antibody (Sigma Aldrich), murine anti-6xHis antibody (Clontech Laboratories, Inc., Mountain View, CA), and goat anti-mouse-PE (Life Technologies) were used in stains at 1 µg/mL, 0.5 µg/mL, and 10

μg/mL, respectively. Chicken IgY polyclonal antibodies were used at 1:100 dilutions, followed by 5 μg/mL goat anti-chicken-PE (EMD Millipore, Billerica, MA).

Yeast-scFv Diversity

The diversity of yeast-scFv selection outputs was determined by BstNI digest. Yeast-scFv plasmid DNA was extracted by phenol-chloroform or with a Zymoprep Yeast Plasmid Miniprep I kit, per manufacturer's instructions (Zymo Research, Irvine, CA). Plasmid was diluted 1:200 in a 50 μL PCR reaction containing 1x Phusion High Fidelity Mastermix (New England Biolabs, Inc., Ipswich, MA), 0.5 μM forward and reverse primers (forward primer: 5'-GTACGAGCTAAAGTACAGTG-3'; reverse primer: 5'-TAGATACCCATACGAGCTTC-'3), and 0.2 M betaine. The reaction mixes were run on the following PCR program: 1) denaturation at 95°C for 5 minutes; 2) 34 amplification cycles of 95°C for 30 seconds, 55°C for 30 seconds, and 72°C for 45 seconds; and 3) final extension at 72°C for 5 minutes.

PCR products were purified by the Qiaquick PCR purification kit (Qiagen, Hilden, Germany). Five hundred ng of purified product were digested with 20 units of BstNI (New England Biolabs, Inc.) at 60°C for 1 hour. The resultant digests were electrophoresed and visualized on 1% agarose.

Yeast-scFv Lyophilization and Stability Study

Selected yeast-scFv were induced in culture, lyophilized, stored dry at 25°C, 37°C, and 45°C for 1 year, and screened month for antigen-binding activity as described[213].

Generation of Nanoyeast-scFv

For experiments based at the University of Queensland, nanoyeast-scFv were generated as previously described[215,216].

For the ELISA experiments performed at the University of Washington, lyophilized yeast-scFv cakes were fragmented by mortar and pestle as before, but they were reconstituted with 2 mL 1x PBS + 5% glycerol + 1 protease inhibitor tablet (Life Sciences) per cake, immediately filtered with no further dilution, and immediately employed in the assays[215,216].

Nanoyeast-scFv Screen Printed Electrode Assay and SERS Assay

Investigation of nanoyeast-scFv performance on gold screen-printed electrodes and in a SERS assay has been previously described[216,217]

Nanoyeast-scFv ELISA

Wells of a Maxisorp flat bottom plate (Nunc) were coated with 50 ng of anti-hemagglutinin rabbit polyclonal antibody (Abcam #ab9110) in 50mM bicarbonate buffer, pH 9.6 overnight at 4°C. The wells were washed three times with 1x PBS + 0.05% Tween 20 (PBS-T) and blocked 90 minutes in 1% BSA in Tris Buffer + 0.05% Tween 20 (TBS-T). Next, wells incubated with 100 µL of 0.5 cake/mL nano yeast-scFv solution (described above) overnight at 4°C.

To confirm nano yeast-scFv binding, a subset of wells received 1 µg of anti-c-myc monoclonal antibody (Santa Cruz Biotechnology, Dallas, TX) and 1:1000 goat anti-mouse HRP (H+L) (Bio-Rad, Hercules, CA) diluted in 1% BSA TBS-T for 1 hour each, with three washes of 1x PBS-T before each step. Meanwhile, antigen binding was tested via titration of EHI_044500 ("Jacob") cognate antigen, followed by 25 ng 1A4 or 1D3 anti-Jacob primary antibody and

1:1000 goat anti-mouse HRP; all reagents were diluted in 1% BSA TBS-T, and each step was one hour incubation at room temperature preceded by three washes in PBS-T. After incubation with goat anti-mouse HRP, all wells were washed 4 times with PBS-T. Wells were developed with 1-step ultra tetramethylbenzidine (TMB) (Life Technologies) for 4 minutes before the reaction was stopped with 2N hydrochloric acid. Absorbance was read at 450 nm.

Supplementary Figures

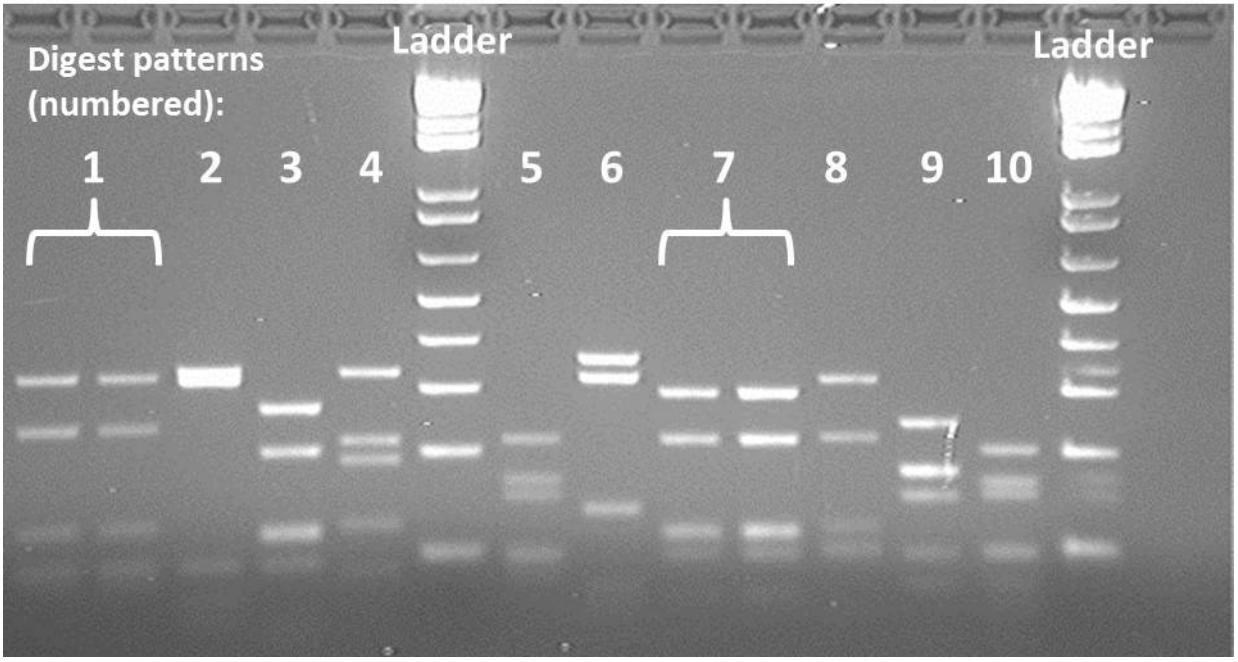

Figure 3S1. The diversity of scFv selected against *Entamoeba histolytica* antigen EHI_033250 as demonstrated by scFv DNA digest with BstNI restriction enzyme. The unique digest patterns observed are numbered 1-10.

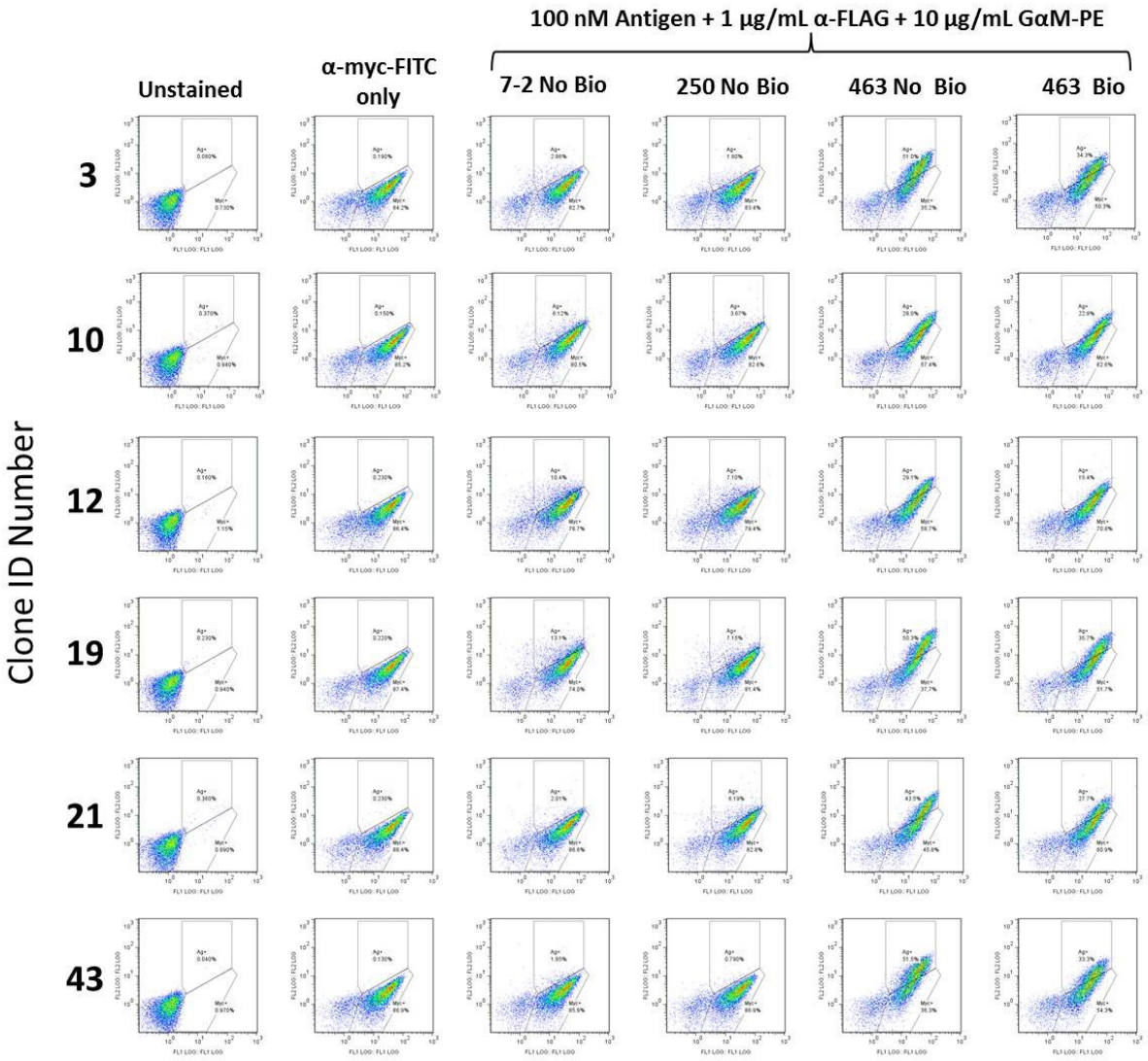

Figure 3S2. Clones selected against EHI_104630 are confirmed to bind non-biotinylated cognate antigen specifically. This panel is not compensated and has individual clones arranged in rows. The "α-myc-FITC only" column confirms each clone expresses scFv and sets the threshold for gating. GαM-PE= R-phycoerythrin-conjugated goat anti-mouse antibody.

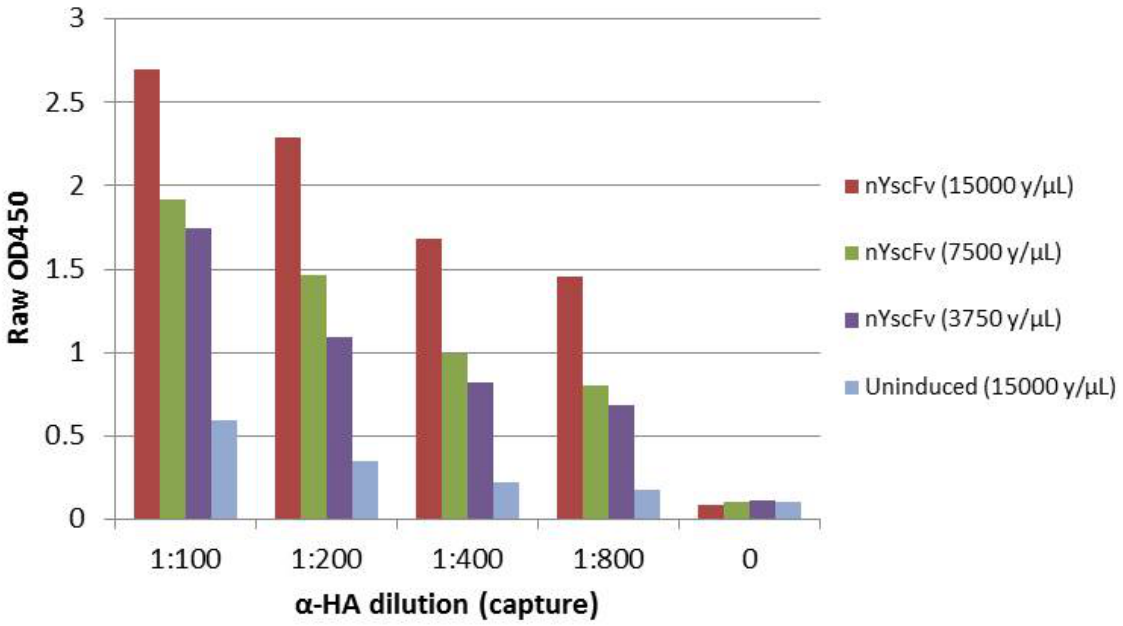

Figure 3S3. Demonstration of nanoyeast-scFv capture in an ELISA format. Nanoyeast-scFv were bound to the plate by coated anti-HA antibody, and detection of the binding was facilitated by anti-c-myc primary antibody with HRP-conjugated secondary. Nanoyeast-scFv reagent was generated at 0.5 cake/mL (15000 y/µL) and titrated by halves. Uninduced nanoyeast-scFv (in which galactose was never added to induce scFv display) was also generated at 0.5 cake/mL.

CHAPTER IV: SPECIES-SPECIFIC IMMUNODETECTION OF AN *ENTAMOEBA HISTOLYTICA* CYST WALL PROTEIN

Species-specific Immunodetection of an *Entamoeba histolytica* Cyst Wall Protein

Lauren J. Spadafora[a], Abdullah Siddique[b], Ibne K. Ali[c,d], Carol A. Gilchrist[c], Tuhinur Arju[b], Benjamin Hoffstrom[e], William A. Petri, Jr.[c], Rashidul Haque[b], and Gerard A. Cangelosi[a,f]

a Department of Global Health, University of Washington, Seattle, WA 98195
b International Centre for Diarrhoeal Disease Research, Bangladesh, Dhaka, Bangladesh 1212
c Department of Medicine, School of Medicine, University of Virginia, Charlottesville, VA 22908
d Current address: Division of Foodborne Waterborne and Environmental Diseases, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30329
e Antibody Technology Core, Fred Hutchinson Cancer Research Center, Seattle, WA 98109
f Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, WA 98195

ABSTRACT

Entamoeba histolytica causes intestinal disease in endemic settings throughout the world. Diagnosis of *E. histolytica* infection would be improved by the identification of biomarkers that are expressed by cysts of *E. histolytica*, but not by cysts of closely related commensal species of *Entamoeba*. Herein, we describe two novel monoclonal antibodies (1A4 and 1D3) produced against a spacer region of the *E. histolytica* Jacob2 lectin, an outer cyst wall protein. These reagents demonstrated no cross-reaction to *E. dispar* recombinant antigen and low picomolar molecular detection limits when paired in ELISA sandwich assays. In an immunofluorescence microscopy assay, the α-Jacob2 murine antibodies labeled cysts of 3 xenically cultured *E. histolytica* isolates but did not label cysts of 3 *E. bangladeshi* isolates. Monoclonal antibody 1A4 did not cross-react with xenic cultures of three *E. dispar* isolates, demonstrating specificity to *E. histolytica,* while monoclonal antibody 1D3 cross-reacted with 2 out of 3 *E. dispar* isolates. Both antibodies labeled cysts in formalin-fixed slides, a potential logistical advantage in some settings. These results demonstrate that antibodies generated against the flexible spacer of *E. histolytica* Jacob2 lectin recognize and bind to Jacob2 protein in whole cysts and are capable of differentiating *Entamoeba* species. Thus, Jacob2 is a promising biomarker for use in diagnosing *E. histolytica* infection.

Keywords: Antibodies, *Entamoeba histolytica*, biomarker, detection, diagnosis, immunofluorescence microscopy, ELISA

Introduction

Entamoeba histolytica is a protozoan parasite that causes an estimated 30-50 million cases of illness and kills 100,000 people annually[104,105]. It has a dual-stage life cycle, consisting of motile trophozoites that colonize and invade the colonic epithelium and of robust, chitinaceous cysts that enter and exit the human body via fecal-oral transmission[111,114]. Infection is often asymptomatic, or it can lead to clinical manifestations that include dysentery, colitis, or extraintestinal abscesses[107,111,114]. Symptoms often resemble other enteric diseases caused by bacteria and viruses, as well as inflammatory bowel disease[111,114]. Control of this organism is particularly important for young children in endemic regions, whose nutrition, growth, and development are negatively impacted by enteric infections and repeated diarrheal episodes[121,225,226].

Detecting *E. histolytica* can be challenging due to the numerous commensal amoeba that colonize humans, some of which look morphologically identical to the pathogen[150,154,227]. Much work has been undertaken to identify and characterize superior biomarkers of infection in stool and of invasive disease in serum and abscess fluid. One such marker is the galactose/N-acetyl galactosamine (Gal/GalNAc) lectin, an adhesion factor that is important to *E. histolytica* trophozoite invasion[123]. This protein is the target of two antigen capture assays that have been widely and successfully used to specifically detect *E. histolytica* infections in fresh stool or liver abscesses[160,180,181]. However, these tests are unable to detect *E. histolytica* cysts in stool or to detect fixed organisms, and the instability of shed, unfixed trophozoites necessitates prompt transport and testing[105,150,158,160]. These limitations may have an impact on *E. histolytica* diagnosis, as one of the tests was found to be at best 79% sensitive relative to the more sensitive qPCR method[161].

A cyst wall lectin with diagnostic potential, named Jacob2, was described recently[193]. It is one of a few proteins known to be expressed only in the *Entamoeba* cyst stage[144,145,228]. In the "wattle and daub" model of *Entamoeba* encystation, the Jacob2 protein first appears in intracellular vesicles and is secreted through the plasma membrane[143,144]. Then, it is tethered by the Gal/GalNAc lectin and binds to chitin to form the cyst wall via three, chitin-binding domains[143–146,193]. Between these domains is a flexible, serine-rich spacer sequence with an amino acid sequence dissimilar between *E. histolytica* and commensal non-pathogenic species *E.dispar*[144,193]. This sequence was also noted to be polymorphic between evaluated *E. histolytica* strains[144,193]. Nevertheless, it was posited that Jacob2 and other cell wall components could be utilized to distinguish *E. histolytica* from similar non-pathogenic species such as *E. dispar*[144,193].

Here, we describe the generation of murine monoclonal antibodies against the variable spacer region of the *E. histolytica* Jacob2 protein. These reagents demonstrated excellent analytical sensitivity in a sandwich ELISA. More importantly, a monoclonal antibody bound *E. histolytica* cysts in xenic culture without cross-reacting to xenic isolates of the commensal species *E. dispar* and of the recently discovered species, *Entamoeba bangladeshi*[227]. This study supports the feasibility of using the Jacob2 spacer region as an *E. histolytica*-specific cyst biomarker.

Results

Generation of Jacob2-specific Antibodies

The target of this study was the Jacob2 lectin, which was previously shown to be expressed only on the cyst cell wall of *Entamoeba* species. The spacer region of this protein was considered potentially useful as a diagnostic target, because its amino acid sequence diverges

between *Entamoeba* species. **Figure 4-1** is a protein sequence alignment of the Jacob2 lectin from *E. histolytica* HM-1:IMSS (EhJacob) and *E. dispar* SAW760 (EdJacob). The highlighted regions were cloned for expression in this study. The full length-proteins have 64% identity and

Figure 4-1. ClustalW alignment of the Jacob2 proteins from *Entamoeba histolytica* HM1:IMSS (EHI_044500) and *Entamoeba dispar* SAW760 (EDI_246160). The ClustalW BLOSUM matrix was used with a gap open cost of 10 and a gap extension cost of 0.1. This figure was generated in Geneious 6.0.3. Cloned residues for "EhJacob" and "EdJacob" recombinant antigens are highlighted in light gray and dark gray respectively

22% gaps; however, the cloned regions (which are mostly comprised of the spacer regions) have 53% identity and 15% gaps. The cloned constructs expressed well in BL21(DE3) cells with codon optimization, and high purity protein was obtained with Ni-NTA affinity chromatography (not shown). The SUMO tag was cleaved from EhJacob protein prior to immunization.

EhJacob protein was combined in a cocktail of four immunogens, and immunization and fusion resulted in 108 hybridoma culture supernatants that were specific to EhJacob by indirect ELISA. This pool was narrowed based on activity and specificity, in a stepwise process summarized in **Table 4-1**. Out of 108 antibodies that bound to EhJacob in the first step, 48 IgG1 and 16 IgG2 antibodies (N = 64) with the highest activity in indirect ELISA (ODs >1.0) were further screened against both EhJacob and EdJacob by using 48 x 16 matrix sandwich ELISAs. Of these, 41 bound EhJacob antigen as part of at least 1 pair. However, only 20 did not cross-react with EdJacob protein. Of these, 11 antibodies were examined in a third step involving further indirect ELISA. Six of the 11 clones were eliminated as SUMO-directed antibodies, based on their cross-reaction with an irrelevant, SUMO-tagged protein. This left 5 antibodies (2 IgG1 and 3 IgG2a) that specifically bound to recombinant EhJacob protein.

TABLE 4-1. SUMMARY OF SCREENS FOR α-JACOB MURINE MONOCLONAL ANTIBODIES

Screening Step	Screening Assay	Binding Properties	Total Screened	Total Positive (%)	By Isotype		
					IgG1	IgG2a	IgG2b
1	Indirect ELISA	Hits specific to EhJacob	1440	108 (7.5%)	78	21	9
2	Sandwich ELISA	Binds to EhJacob	64	41 (64.1%)	29	11	1
		Not Cross-Reactive to EdJacob		20 (31.3%)	9	11	0
3	Indirect ELISA	Not Cross-Reactive to SUMO	11	5 (45.5%)	2	3	0

Limit of Detection in Sandwich Assays

Out of the 5 EhJacob-specific antibodies, 1 IgG1 antibody (designated 1A4) and 1 IgG2a antibody (designated 1D3) were chosen for further investigation and purified. Their limit of detection as a pair in a sandwich ELISA was determined (**Figure 4-2**). There was no signal when 1A4 capture antibody was eliminated from the sandwich, indicating that the Jacob antigens did not non-specifically bind to the plate (data not shown). With 1A4 as the capturing antibody, the two antibodies could detect recombinant *E. histolytica* Jacob2 antigen in buffer at concentrations down to 9.8 pM (67 pg). The signal at this antigen concentration was more than three standard deviations greater than the signal with non-cognate *Entamoeba dispar* Jacob2 at 2500 pM concentration.

Microscopy

Although the 1A4 and 1D3 antibodies exhibited excellent analytical sensitivity and

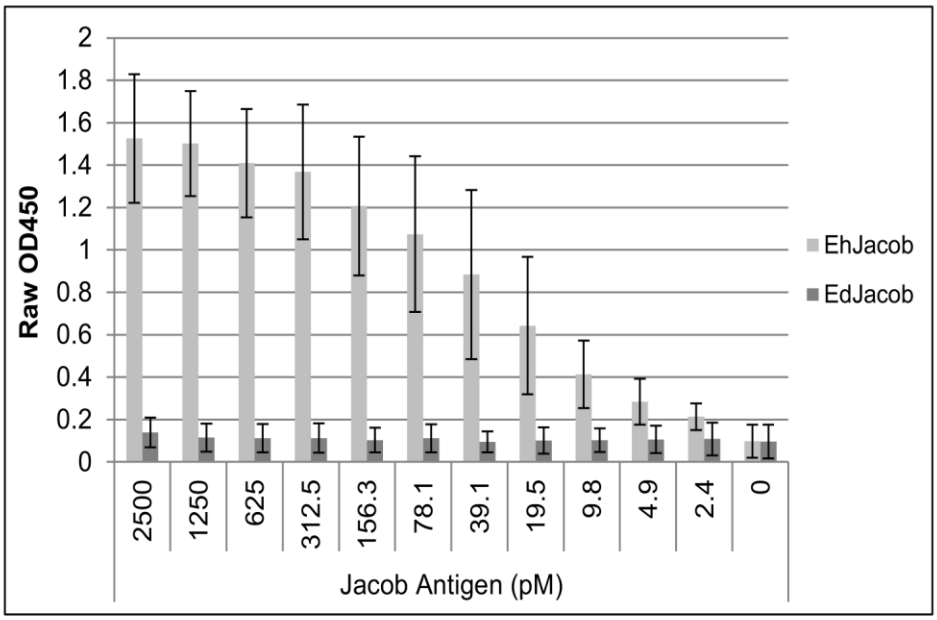

Figure 4-2. Monoclonal antibodies 1A4 and 1D3 detecting *Entamoeba histolytica* and *Entamoeba dispar* recombinant Jacob2 antigen in a sandwich ELISA. Measured antigen concentration ranged from 2.4-2500 pM. Columns and error bars represent mean OD450 ± standard deviation of three replicate assays. Limit of detection (LOD) was calculated as 3 standard deviations above the mean OD450 of EdJacob at 2500 pM concentration.

selectivity in an ELISA format, it is not known whether Jacob2 can be detected as a free soluble protein in patient stool samples. Therefore, we evaluated whether these antibodies can detect whole *Entamoeba* cysts in immunofluorescence microscopy assays conducted on xenic cultures of *E. histolytica*, *E. dispar,* and *E. bangladeshi*. Three independent isolates derived from stool were tested for each species. Xenic culture samples were smeared onto slides, fixed, blocked, and stained with 1A4 or 1D3 antibody, followed by labeling with goat anti-mouse Alexa Fluor 488 secondary. Calcofluor White M2R stain was used as a marker of encystation, as previously established [138]. **Figure 4-3** is a panel of representative photos for each species with each antibody. By eye, IgG1 antibody 1A4 only produced fluorescent cyst-like objects when staining *E. histolytica* isolates. In contrast, cross-reaction was seen when Ig2a antibody 1D3 stained *E. dispar* isolates.

Computer-assisted image analysis was used to quantify complete observations. "Cyst-like" objects were identified post-photo acquisition based on Calcofluor intensity, size, and circularity. Mean Alexa Fluor 488 fluorescence intensities of identified "cyst-like" objects were then calculated. Results are shown in **Figure 4-4** by species and isolate. When measured by a Mann Whitney U hypothesis test, the increased fluorescence of 1A4-stained *E. histolytica* cyst objects over 1A4-stained *E. dispar* and *E. bangladeshi* cyst objects was statistically significant (one-tailed p-value=0.025) (**Figure 4-4A**). The higher fluorescence of 1D3 stained *E. histolytica* objects compared to 1D3-stained *E. bangladeshi* was also statistically significant (one-tailed p=0.025), but the fluorescence of *E. histolytica* and *E. dispar* objects was not significantly different (one-tailed p =0.26) (**Figure 4-4B**). Two of the three *E. dispar* isolates appeared to cross-react with this antibody.

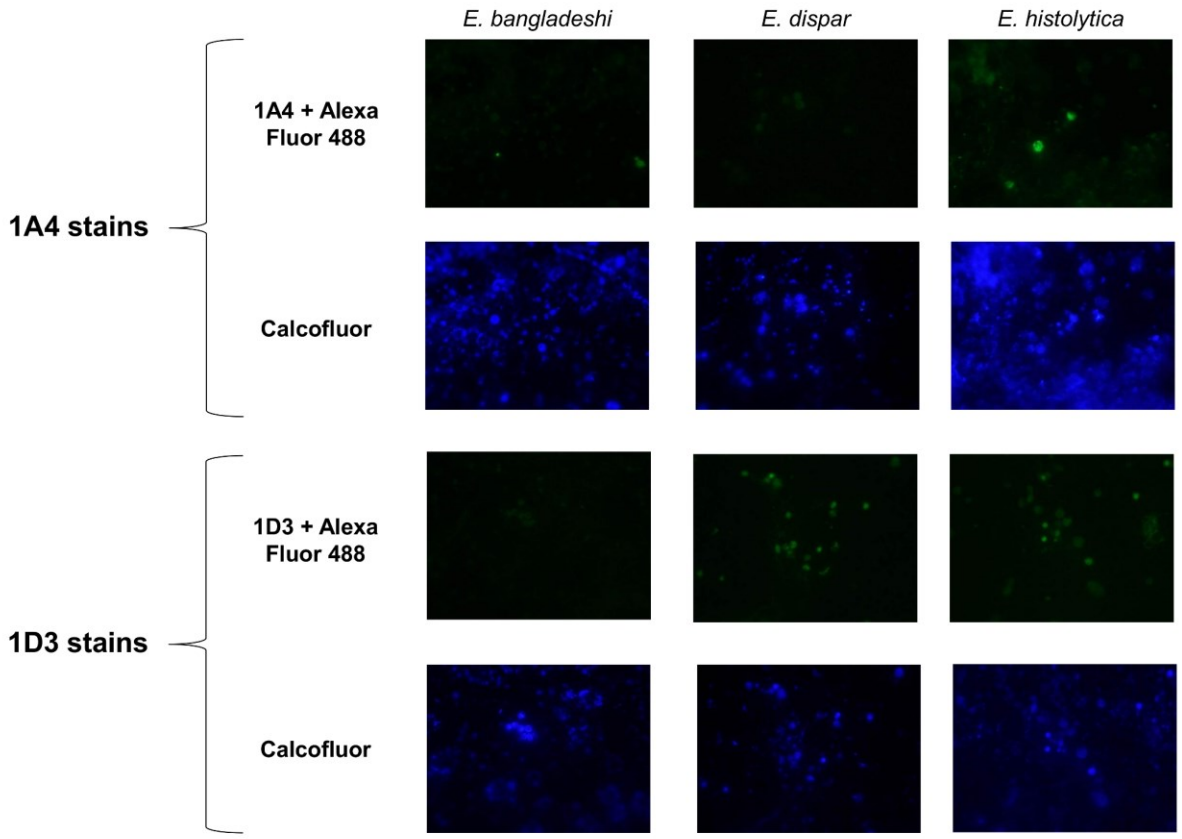

Figure 4-3. Representative photos from an anti-Jacob2 immunofluorescence microscopy assay. Isolates were doubled stained with 0.1% Calcofluor White M2R and primary anti-Jacob2 monoclonal antibody (1A4 or 1D3) with goat anti-mouse Alexa Fluor 488.The three species examined were pathogen *Entamoeba histolytica* (1st column) and commensals *Entamoeba dispar* and *Entamoeba bangladeshi* (2nd and 3rd columns). Calcofluor was utilized to identify chitinaceous *Entamoeba* cysts (2nd and 4th rows).

These quantitative results supported the results seen by eye. Although many cyst-like objects did not stain with either antibody regardless of species, strongly-staining objects were confined to *E. histolytica* in the case of 1A4 or to *E. histolytica* + *E. dispar* in the case of 1D3.

Discussion

This report describes two novel antibodies that detect *Entamoeba* cysts. They were generated against the flexible, serine-rich spacer of the Jacob2 lectin, expressed as a recombinant protein fragment in the absence of more conserved portions of Jacob2. Both antibodies were

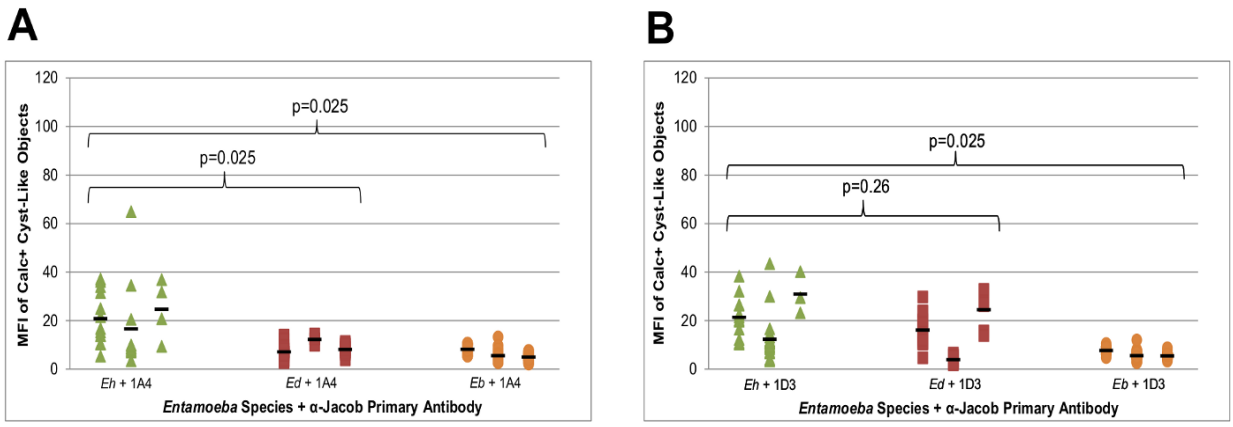

Figure 4-4. *Entamoeba* species specificities of anti-Jacob antibodies 1A4 (A) and 1D3 (B) in immunofluorescence microscopy. Smears of xenic *Entamoeba* isolates were doubly stained with 0.1% Calcofluor White M2R and an anti-Jacob primary antibody with goat anti-mouse Alexa-Fluor 488. Each data point represents the mean Alexa-Fluor 488 intensity of a Calcofluor-positive, cyst-like object detected by microscopy. The data points are arranged horizontally by isolate, and black horizontal bars represent the mean Alexa-Fluor 488 intensity of each isolate. The overall mean Alexa-Fluor 488 intensities of the *Entamoeba histolytica* isolates (n=3) were compared to those of *E. dispar* isolates (n=3) and of *Entamoeba bangladeshi* isolates (n=3) with the Mann Whitney U test.

specific to *E. histolytica* recombinant antigen, and they detected this antigen with excellent sensitivity in a sandwich ELISA format. However, only the IgG1 antibody designated 1A4 bound whole *E. histolytica* cysts in culture without cross-reacting either to *E. dispar* or *E. bangladeshi* cysts. The reason 1D3 was cross-reactive with *E. dispar* when applied to whole cysts (**Figures 4-3 & 4-4**), but not when applied to recombinant EdJacob (**Table 4-1**), is not known. It may reflect different physical conditions and availabilities of epitopes in the different assay formats. There could also be a cross-reactive epitope in the native antigen of some *E. dispar* isolates. Protein sequence analysis conducted on diverse strains could help resolve this question.

The *E. histolytica* Jacob2 lectin drew interest for its spacer divergent from *E. dispar* and for its localization on the cyst cell wall [143,144,193]. In this study, the spacer of interest was expressed alone to increase success in finding a species-specific reagent. However, as with other *E. histolytica* surface antigens, the spacer has been found to vary by tandem repeats amongst *E. histolytica* clinical isolates, and such variation may possibly be found within close phylogenetic

groupings [229,230]. Hence, further investigation with more *Entamoeba* strains must be conducted, given the small and geographically limited sample used in this study. Additionally, there are at least three additional murine antibodies (2C7, 4A4, 7D6) and three yeast-displayed, human single chain fragment variable (scFv) antibodies that could be further investigated for desired research and diagnostic applications.

.This investigation demonstrated detection of cysts in formalin-fixed samples. Many currently available and reported assays for *E. histolytica* detection detect trophozoites only, or are unable to detect organisms in fixed samples[115,150]. As trophozoites are unstable outside of the host, this has required that testing be performed on fresh samples that have been quickly transported to a laboratory [105,150]. Utilization of 1A4 or 1D3 could allow for detection of *E. histolytica* in stool samples that were fixed upon collection and tested at a later time. This may be a significant logistical advantage in some settings.

Despite limitations, this work shows promise toward validating a new biomarker for use in species-specific detection of *E. histolytica* cysts. Monoclonal antibody 1A4 is one of few cyst-targeted immunoreagents to distinguish *E. histolytica* from nonpathogenic *Entamoeba* species, and, to our knowledge, the only reported reagent to bind fixed *E. histolytica* cysts specifically[115]. These findings may ultimately benefit end-users in diagnosis of *E. histolytica* infection and in reducing the global burden of diarrheal disease caused by this pathogen.

Acknowledgements

We would like to thank Meagan Ashworth Dirac and Annie Becker for human subjects compliance support, and Moira Kearney, Jingwen Xiao, and Kris Weigel for helpful input.

Competing Interests

WAP receives licensing income from TechLab Inc. for amebiasis diagnostics. These are donated in their entirety to the American Society of Tropical Medicine and Hygiene without benefit to Dr. Petri.

Funding Sources

This work was supported by NIH grants U01AI082186 to GAC and 5R01 AI043596 to WAP. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Methods

Ethics Statement

Entamoeba cysts were cultured from stool samples collected with approval from the Ethics Review Committee of the International Centre for Diarrheal Disease Research, Bangladesh (ICDDR,B) and the Institutional Review Board (IRB) of the University of Washington. Informed, written consent was obtained from patients or from guardians of subjects ages 1-17, and assent was obtained from subjects ages 8-17. Monoclonal antibody production was approved by the Fred Hutchinson Cancer Research Center Institutional Animal Care and Use Committee (IACUC).

Bioinformatics

The Jacob2 sequences for *E. histolytica* strain HM1:IMSS (EHI_044500) and for *E. dispar* strain SAW760 (EDI_246160) were stored on Geneious version 6.0.3 and aligned

utilizing the ClustalW BLOSUM cost matrix, with a gap open cost of 10 and a gap extend cost of 0.1.

Recombinant Protein Production

All cloning and expression reagents were obtained from Life Technologies (Carlsbad, CA) unless otherwise noted. Residues 159-481 of the *E. histolytica* strain HM-1:IMSS Jacob2 protein (EhJacob) were codon-optimized and TOPO-TA-cloned into the pET SUMO vector to produce recombinant proteins with an N-terminal 6xHis-SUMO tag[197]. Residues 212-560 of the *E. dispar* strain SAW670 (EdJacob) were cloned into the Gateway pDEST17 vector with an N-terminal 6xHis tag. The constructs were expressed in BL21(DE3) *Escherichia coli* cells, and protein was purified via Ni-NTA resin (Qiagen, Hilden, Germany). The SUMO tag on the EhJacob protein was cleaved with SUMO protease (Lifesensors, Malvern, PA) and removed by Ni-NTA resin according to kit instructions.

Murine Monoclonal Antibody Production

Murine monoclonal antibodies were generated by the Antibody Technology Core at the Fred Hutchinson Cancer Research Center (Seattle, WA). EhJacob was dialyzed into 1x PBS and combined with three additional *E. histolytica* recombinant antigens (EHI_101240, EHI_070730, and EHI_104630) for immunization. The multiplex antigen formulation was used to generate antibodies for multiple investigations including this one. Three sets of splenic cell fusions were performed: the first was from mice with the highest titer to antigen and was screened for specificity via an indirect ELISA, while the second and third were from mice with lower titers to antigen and were first selected for IgG secretion with a ClonePixII system (Molecular Devices,

Sunnyvale, CA), followed by target specific indirect ELISA screening. The isotypes were determined by indirect ELISA assays.

Screening of Monoclonal Antibodies by Indirect and Matrix Sandwich ELISAs

The 48 IgG1 and 16 IgG2 antibodies with highest binding activity (OD >1.0) were screened as pairs in 48 x 16 matrix sandwich ELISAs. First, anti-IgG1 antibody (0.5 ug/mL, Life Technologies A10538) was coated overnight at 4°C to high binding, 384-well plates (Greiner Bio-One, Kremsmünster, Austria). Next, the following materials were incubated for 1 hour at room temperature with subsequent 3 rounds of washing with PBS + 0.05% Tween 20 (PBS-T): 1) 25 µL of 1:8 IgG1 supernatants; 2) 125 ng/mL EhJacob (with SUMO tag) or EdJacob antigen; 3) 25 µL of 1:8 IgG2 supernatants; and 4) 1:2000 HRP-conjugated anti-IgG2a or –IgG2b antibodies (Life Technologies A10685 and M32507). Wells were developed with 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid) (ABTS) (Kirkegaard & Perry Laboratories, Gaithersburg, MD) for 15 minutes and read at 405 nm, with threshold set at an optical density of 0.7.

The specificities of top antibodies were checked once again by indirect ELISA, this time incorporating two non-cognate antigens, one of which had a SUMO tag. Wells of a Nunc Maxisorp flat bottom plate (Nunc, Rochester, NY) were coated with 200 ng of antigen diluted in 50 mM bicarbonate buffer, pH 9.6 (Sigma Aldrich, St. Louis, MO) overnight at 4°C. After the plate was washed 3 times in PBS-T and blocked with Starting Block (Life Technologies) for 1 hour at room temperature, wells incubated with selected hybridoma supernatants diluted 1:5-1:20 in Starting Block for 1 hour at room temperature. The plate was washed again and incubated with 1:1000 of goat anti-mouse HRP (H+L) (Bio-Rad Laboratories, Inc., Heracles, CA) for 1

hour at room temperature. One-step Ultra TMB substrate (Life Technologies) was added after 4 additional washes with PBS-T, and the reaction was stopped with 2N hydrochloric acid upon development of color. The microplates were read at 450 nm.

Sandwich ELISA with Purified Antibody

Wells in flat-bottom Maxisorp microplates were coated with 0.15 μg of purified monoclonal antibody 1A4 diluted in 100mM bicarbonate buffer, pH 9.6 (Sigma Aldrich), overnight at 4°C. Simultaneously, an additional set of wells were coated with blank buffer to serve as a control with no capture antibody. After 3 washes in PBS-T, wells were blocked with 1% BSA in PBS-T for 90 minutes at room temperature. Next, a serial titration of SUMO-EhJacob, ranging 16.4 pg to 17.1 ng, was plated and incubated for 60 minutes at room temperature, followed by 3 washes in PBS-T. Monoclonal antibody 1D3 (1.2 ug/mL) and 125 ng/mL HRP-conjugated goat anti-mouse IgG2a (#ab97245, Abcam, Cambridge, MA) incubated for 1 hour at RT in succession, followed by three and four washes in PBS-T respectively. Finally, wells were developed with 100 μL of 1-Step Ultra TMB solution for 60-80 seconds and stopped with 2N hydrochloric acid. The microplates were read at 450 nm.

Xenic Entamoeba *spp. culture in Robinsons Diphasic Media*

Stool samples were obtained from children enrolled in an ongoing community-based prospective cohort study of enteric infections [231]. Diarrheal and monthly surveillance stools were examined by microscopy, and a 50 mg aliquot of any prospective *Entamoeba* positive feces was inoculated into a 7 mL glass McCartney bottle containing a sterile agar slant and a liquid overlay of Robinson's media, bacto-peptone, erythromycin, and 10 mg of rice starch [232,233]. Bottles were incubated at 37°C for 24 hours. Next, a drop of culture from the fecal-starch layer was drawn up

and examined by microscopy. If no amoeba were found, additional rice starch was added to the bottle, and the culture was incubated for an additional 48 hours before re-examination. If amoeba were identified, the culture was passaged every 48 hours by inoculation of 0.1 mL of fecal-starch layer into a fresh bottle of media. *Entamoeba dispar* and *E. bangladeshi* isolates were identified by PCR, whereas *E. histolytica isolates* were confirmed by both PCR and ELISA (Techlab, Blacksburg, VA).

Immunofluorescence Assays

Xenic cultures were briefly checked for the presence of *Entamoeba*-like organisms by light microscopy prior to slide preparation. Smears were first air dried onto Merifluor treated slides (Meridian Bioscience, Inc., Cincinnati, OH) and then fixed in 10% neutral buffered formalin and in 100% methanol (Merck, Darmstadt, Germany) for 1 minute each. After quick washes in PBS and PBS-T, they were blocked in 1% BSA PBS-T for 1-2 minutes at room temperature. Next, the smears were stained with 2.9 µg/mL 1A4 or 2.3 µg/mL 1D3 purified monoclonal for 1 hour and with 20 µg/mL goat anti-mouse Alexa Flour 488 (Life Technologies) for 30 minutes in a humid chamber at room temperature, with a brief PBS-T wash after each incubation. Finally, the smears were stained with 0.1% Calcofluor White M2R (Sigma Aldrich) for 5 minutes at room temperature in the dark. The smears were examined on an Olympus BX53 microscope under 400x magnification, and photos were acquired with an Olympus DP21 camera. Photos were taken under a UV filter at 10 millisecond exposure time and under a FITC filter at 167 millisecond exposure time.

Image & Statistical Analyses

Each photo taken under the UV filter was sized to 1300 x 975 pixels on GIMP2.8 and converted to 8-bit grayscale on ImageJ (National Institutes of Health) [234]. Background was subtracted with a 50-pixel rolling ball radius, sliding paraboloid enabled and smoothing disabled. The image was then converted to binary after setting the threshold to ≥12, and the watershed algorithm was applied to separate clustered particles. Resultant particles that were 1000-5000 square pixels in area and with circularity ranging 0.65-1.00 were labeled as "cyst-like" objects. Labels on the UV filter photos were transferred to the corresponding FITC filter photos, and the mean "gray value" for each labeled object was obtained.

Statistical analysis was conducted with Microsoft Excel and the Real Statistics Resource Pack software (Release 3.5). The mean gray values of the FITC filter photos, a proxy for Alexa Fluor 488 fluorescence intensity, were grouped by isolate. The overall mean fluorescence intensities of the *E. histolytica* isolates were compared to *E. dispar* isolates and *E. bangladeshi* isolates with the Mann Whitney U test ($n=m=3$, $\alpha=0.05$).

CHAPTER V: CONCLUSIONS

I. LEVERAGING ALTERNATIVE AFFINITY REAGENTS

The first aim of this work was to develop an affinity reagent pipeline in which alternative affinity reagents can be generated more quickly and affordably relative to classical monoclonal antibodies. Such an effort has become a critical research focus as proteomics enters the spotlight in the post-genomic era. Several similar endeavors are underway worldwide, collectively utilizing many affinity reagent scaffolds and selection methodologies[2]. It is predicted that not one class of affinity reagent will emerge as the most useful and successful, but rather, different classes and methods of their generation will be employed for particular applications[2].

In contrast to many engineered affinity reagents, our strategy focuses on the power of selection. It uses pre-constructed, displayed affinity reagents in the environment in which they were selected, attached to the yeast cell wall. Extensive effort has been dedicated to optimizing the yield of soluble, recombinant antibody that had been isolated from a display library[97,102]. It was noted that, generally, higher display of protein on a yeast display library correlated with higher thermal stability and secretion yield[101]. However, due to soluble scFv misfolding and inactivation, some previous endeavors had to select and characterize at least three yeast-scFv clones for every specific and functional soluble scFv[102,212]. Our investigation circumvents this additional labor and succeeding failure with the generation and validation of nanoyeast-scFv, a novel affinity reagent in which displayed scFv are maintained on fragmented yeast cell. It has now been demonstrated that nanoyeast-scFv, derived from three separate yeast-scFv clones, remain functional when tethered to a variety of solid support assays, with limits of detection in nanograms or picograms per milliliter in buffer containing complex biological substance[215–217].

As development of nanoyeast-scFv continues, many questions must be answered:

1) **What is the structure of nanoyeast-scFv?** It is especially critical to know what remnants of the yeast cell wall are present and influencing the functionality of the scFv. This knowledge will be important as our consortium continues to produce nanoyeast-scFv and improve upon their preparation. Sizing by filtration, electron microscopy, and Western blotting have already been undertaken[235], and experiments utilizing mass spectrometry and perhaps crystallography should be conducted to obtain a more complete understanding of these reagents.
2) **Can nanoyeast-scFv be utilized beyond solid support assays?** This work has only investigated assays in which nanoyeast-scFv are tethered to a solid support (electrode surface, polystyrene well) via anti-HA epitope tag antibody. To become a viable alternative to the traditional monoclonal antibody, nanoyeast-scFv will also need to function as fully soluble reagents in assays such as Western blot, immunofluorescence microscopy, and as a signal reagent in ELISA or flow cytometry. Methods to enrich and elute fresh nanoyeast-scFv will need to be developed.
3) **How can nanoyeast-scFv generation be standardized and optimized?** We have yet to determine whether every functional yeast-scFv clone will result in a functional nanoyeast-scFv. Moreover, methods to produce these reagents remain crude and are likely to introduce variability between lots of reagent. Factors such as operator's technique, time and thoroughness of fragmentation, and room conditions could affect output. Methods of maintaining the conditions between lots (i.e. an automatic milling procedure) should be developed.

It remains to be seen whether phage or bacteria display systems can be leveraged in a similar fashion. Out of the display systems available, heterologous yeast display was the best choice for this project. As eukaryotes, yeast species have a higher success rate of correctly folding and displaying mammalian-derived proteins, such as antibodies[81]. Mammalian cell display has been employed, but yeast is safer, less expensive to maintain, and much easier to culture[97]. The yeast display selection process is straightforward, requiring simple magnetic selections prior to FACS enrichment, and fast, requiring only two to three weeks to complete[82,208]. The yeast-scFv selection success rate for this project was fair (47% of antigens with specific yeast-scFv), but antigen characteristics (i.e. lack of immunogenicity, biotinylation) were a main contribution to selection failure. Additionally, there are other aspects of yeast display that were not often utilized in this project: scFv biophysical characterizations on the yeast cell wall (affinity, stability) and *in vitro* affinity maturation. Future incorporation of these techniques could inform the selection and ultimate generation of superior nanoyeast-scFv reagents.

II. INVESTIGATION OF *ENTAMOEBA HISTOLYTICA* CYST BIOMARKERS

The secondary aim of this work was to validate a novel *Entamoeba histolytica* biomarker for improved diagnosis in stool. The project's strategy was to target *E. histolytica* cysts, which are more frequent and stable in stool yet are not the target of many commercial diagnostic immunoassays[105,150,158]. The goal was to identify a protein target that was more sensitive than the Gal/GalNAc lectin, a trophozoite surface protein with 79% sensitivity, and that was non-crossreactive to nonpathogenic *Entamoeba* species[161,171].

Many previous attempts to develop *E. histolytica*-specific affinity reagents used the following strategy: 1) immunize an animal with whole or lysed laboratory trophozoites; 2) identify monoclonal antibodies that can differentiate between pathogen and non-pathogen; and 3) attempt to identify the biomarker that produced such antibodies via cDNA screening and Western blots[165,170,173]. Unfortunately, *E. histolytica* cysts generally cannot be obtained in large quantities *in vitro*. Only one study producing cyst-specific monoclonal antibodies was able to do so by fortuitously obtaining a single stool sample that contained no other parasitic species and from which enough cysts for immunization could be purified. In this work, we instead selected individual candidate biomarkers from proteomic and transcriptomic data obtained from *E. histolytica* cysts[138,186]. Utilizing a list of proteomic targets, we narrowed the list by investigating targets that were in a majority of tested fecal samples and/or were not down-regulated during encystation. This rational selection of biomarkers produced a list of 54 potential targets.

Attrition was seen in the recombinant production of these biomarkers and in subjecting them to yeast-scFv selections- out of 54 targets, we have only produced yeast-scFv against seven (13%). However, this is still the largest endeavor undertaken to select alternative affinity reagents against *E. histolytica*, as other studies have selected against laboratory trophozoites, as described above, or focused on a single target[192,220,222,223,236]. The success of this study likely could have been improved with large quantities of native *E. histolytica* cysts to use as selection material. However, until a way to obtain such cysts is provided by *in vitro* methods or recombinant eukaryotic expression is streamlined, the strategy of our pipeline is the best available[237].

We have yet to screen our nanoyeast-scFv reagents against clinical *E. histolytica* samples, but it is our hope we will find at least one reagent that can bind its native target and do so with

superior sensitivity. In the meantime, we concurrently sought production of traditional murine monoclonal antibodies against our selected *E. histolytica* biomarkers. Antibodies produced against one target, the Jacob2 cyst wall lectin, demonstrated species-specific cyst binding when screened with a small subset of fixed, xenic cultures. With further and geographically-expanded sampling and screening, these reagents' ability to consistently detect a polymorphic target and the target's specificity and sensitivity will be determined[193]. This endeavor is the second known attempt to produce *E. histolytica* cyst-specific antibodies and the first to produce reagents that can bind such cysts in fixative[115]. The ability to detect fixed organisms will significantly improve *Entamoeba* detection by way of easier sample collection and handling.

III. FINAL THOUGHTS

Overall, this study has made important contributions to the fields of affinity reagent development and *E. histolytica* diagnosis. As nanoyeast-scFv reagents continue development, their utilization can be employed toward detection of many, other infectious diseases and of human targets critical in chronic and non-communicable disorders. Any endeavors with such reagents will have superior turnaround time and cost, allowing researchers to address human illnesses with the urgency required.

BIBLIOGRAPHY

1. Uhlén, M. Affinity as a tool in life science. *Biotechniques* **44,** 649–54 (2008).
2. Marx, V. Calling the next generation of affinity reagents. *Nat. Methods* **10,** 829–833 (2013).
3. Stoevesandt, O. & Taussig, M. J. Affinity reagent resources for human proteome detection: initiatives and perspectives. *Proteomics* **7,** 2738–50 (2007).
4. Bradbury, A. R. M., Sidhu, S., Dübel, S. & McCafferty, J. Beyond natural antibodies: the power of *in vitro* display technologies. *Nat. Biotechnol.* **29,** 245–254 (2011).
5. Sidhu, S. S. Antibodies for all: the case for genome-wide affinity reagents. *FEBS Letters* **586,** 2778–2779 (2012).
6. Lipman, N. S., Jackson, L. R., Trudel, L. J. & Weis-Garcia, F. Monoclonal versus polyclonal antibodies: distinguishing characteristics, applications, and information resources. *ILAR J.* **46,** 258–268 (2005).
7. Bjørneboe, M. & Gormsen, H. Experimental studies on the role of plasma cells as antibody producers. 1942. *APMIS* **115,** 440–483; discussion 484–485 (2007).
8. Bjørneboe, M., Gormsen, H. & Lundquist, F. Further experimental studies on the role of the plasma cells as antibody producers. *J. Immunol.* **55,** 121–129 (1947).
9. Edelman, G. M. *et al.* The covalent structure of an entire gammaG immunoglobulin molecule. *Proc. Natl. Acad. Sci. U. S. A.* **63,** 78–85 (1969).
10. Edelman, G. M. & Gall, W. E. The antibody problem. *Annu. Rev. Biochem.* **38,** 415–466 (1969).
11. Llewelyn, M. B., Hawkins, R. E. & Russell, S. J. Discovery of antibodies. *BMJ* **305,** 1269–1272 (1992).
12. Wu, T. T. & Kabat, E. A. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J. Exp. Med.* **132,** 211–250 (1970).
13. Kabat, E. A. & Wu, T. T. Attempts to locate complementarity-determining residues in the variable positions of light and heavy chains. *Ann. N. Y. Acad. Sci.* **190,** 382–393 (1971).
14. Hoogenboom, H. R. Selecting and screening recombinant antibody libraries. *Nat. Biotechnol.* **23,** 1105–16 (2005).
15. Tonegawa, S. Somatic generation of antibody diversity. *Nature* **302,** 575–581 (1983).
16. Early, P., Huang, H., Davis, M., Calame, K. & Hood, L. An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: VH, D and JH. *Cell* **19,** 981–992 (1980).
17. Tomlinson, I. M., Walter, G., Marks, J. D., Llewelyn, M. B. & Winter, G. The repertoire of human germline VH sequences reveals about fifty groups of VH segments with different hypervariable loops. *J. Mol. Biol.* **227,** 776–798 (1992).
18. Ravetch, J. V, Siebenlist, U., Korsmeyer, S., Waldmann, T. & Leder, P. Structure of the human immunoglobulin mu locus: characterization of embryonic and rearranged J and D genes. *Cell* **27,** 583–591 (1981).
19. Corbett, S. J., Tomlinson, I. M., Sonnhammer, E. L., Buck, D. & Winter, G. Sequence of the human immunoglobulin diversity (D) segment locus: a systematic analysis provides no evidence for the use of DIR segments, inverted D segments, ‘minor’ D segments or D-D recombination. *J. Mol. Biol.* **270,** 587–597 (1997).

20. Matsuda, F. *et al.* The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. *J. Exp. Med.* **188,** 2151–2162 (1998).
21. Zachau, H. G. The immunoglobulin kappa locus-or-what has been learned from looking closely at one-tenth of a percent of the human genome. *Gene* **135,** 167–173 (1993).
22. Vasicek, T. J. & Leder, P. Structure and expression of the human immunoglobulin lambda genes. *J. Exp. Med.* **172,** 609–620 (1990).
23. Williams, S. C. *et al.* Sequence and evolution of the human germline V lambda repertoire. *J. Mol. Biol.* **264,** 220–232 (1996).
24. Kawasaki, K. *et al.* One-megabase sequence analysis of the human immunoglobulin lambda gene locus. *Genome Res.* **7,** 250–261 (1997).
25. Dreyer, W. J. & Bennett, J. C. The molecular basis of antibody formation: a paradox. *Proc. Natl. Acad. Sci. U. S. A.* **54,** 864–869 (1965).
26. Brack, C., Hirama, M., Lenhard-Schuller, R. & Tonegawa, S. A complete immunoglobulin gene is created by somatic recombination. *Cell* **15,** 1–14 (1978).
27. Bernard, O., Hozumi, N. & Tonegawa, S. Sequences of mouse immunoglobulin light chain genes before and after somatic changes. *Cell* **15,** 1133–1144 (1978).
28. Alt, F. W. & Baltimore, D. Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-JH fusions. *Proc. Natl. Acad. Sci. U. S. A.* **79,** 4118–4122 (1982).
29. Eisen, H. N. & Siskind, G. W. Variations in affinities of antibodies during the immune reponse. *Biochemistry* **3,** 996–1008 (1964).
30. Berek, C. & Milstein, C. Mutation drift and repertoire shift in the maturation of the immune response. *Immunol. Rev.* **96,** 23–41 (1987).
31. Behring, E. von & Kitasato, S. Ueber das Zustandekommen der Diphtherie-Immunität und der Tetanus-Immunität bei Thieren. *Dtsch. Med. Wschr* **16,** 1113–1114 (1890).
32. Winau, F. & Winau, R. Emil von Behring and serum therapy. *Microbes Infect.* **4,** 185–188 (2002).
33. Ehrlich, P. Croonian Lecture: On Immunity with Special Reference to Cell Life. *Proc. R. Soc. London* **66,** 424–448 (1899).
34. Winau, F., Westphal, O. & Winau, R. Paul Ehrlich--in search of the magic bullet. *Microbes Infect.* **6,** 786–9 (2004).
35. Lesky, E. Viennese serological research about the year 1900: its contribution to the development of clinical medicine. *Bull. N. Y. Acad. Med.* **49,** 100–111 (1973).
36. Gruber, M. & Durham, H. E. Eine neue Methode zur raschen Erkennung des Choleravibrio und des Typhusbacillus. *Münchener Med. Wochenschrift* **43,** 285–286 (1896).
37. Kraus, R. Ueber specifische reaktionen in keimfreien filtraten aus Cholera, Typhus, U. Pestvouillon Culturen, erzeugt durch homologes serum. *Wiener Klin Wochenschr* **32,** 736 (1897).
38. Landsteiner, K. *The Specificity of Serological Reactions*. (Charles C. Thomas, Springfield, Illinois, USA 1936).
39. Pauling, L., Campbell, D. & Pressman, D. The nature of the forces between antigen and antibody and of the precipitation reaction. *Physiol. Rev.* **23,** (1943).
40. Heidelberger, M., Goebel, W. F. & Avery, O. T. The soluble specific substance of pneumococcus : third paper. *J. Exp. Med.* **42,** 727–745 (1925).

41. Felton, L. D. Concentration of pneumococcus antibody. *J. Infect. Dis.* **43,** 543–553 (1928).
42. Towbin, H., Staehelint, T. & Gordon, J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets : procedure and some applications. **76,** 4350–4354 (1979).
43. Engvall, E. & Perlmann, P. Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G. *Immunochemistry* **8,** 871–874 (1971).
44. Coons, A. H., Creech, H. J. & Jones, R. N. Immunological properties of an antibody containing a fluorescent group. *Proc. Soc. Exp. Biol. Med.* **47,** 200–202 (1941).
45. Coons, A. H., Creech, H. J., Jones, R. N. & Berliner, E. The demonstration of pneumococcal antigen in tissues by the use of fluorescent antibody. *J. Immunol.* **45,** 159–170 (1942).
46. Lanier, L. L. & Warner, N. L. Paraformaldehyde fixation of hematopoietic cells for quantitative flow cytometry (FACS) analysis. *J. Immunol. Methods* **47,** 25–30 (1981).
47. Hjelm, B. *et al.* Generation of monospecific antibodies based on affinity capture of polyclonal antibodies. *Protein Sci.* **20,** 1824–1835 (2011).
48. Payne, W. J., Marshall, D. L., Shockley, R. K. & Martin, W. J. Clinical laboratory applications of monoclonal antibodies. *Clinical Microbiology Reviews* **1,** 313–329 (1988).
49. Marx, V. Finding the right antibody for the job. *Nat. Methods* **10,** 703–707 (2013).
50. Kohler, G. & Milstein, C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* (1975).
51. Mirza, I. H., Wilkin, T. J., Cantarini, M. & Moore, K. A comparison of spleen and lymph node cells as fusion partners for the raising of monoclonal antibodies after different routes of immunisation. *J. Immunol. Methods* **105,** 235–243 (1987).
52. Bordeaux, J. *et al.* Antibody validation. *BioTechniques* **48,** 197–209 (2010).
53. Spicer, S. S., Spivey, M. A., Ito, M. & Schulte, B. A. Some ascites monoclonal antibody preparations contain contaminants that bind to selected Golgi zones or mast cells. *J. Histochem. Cytochem.* **42,** 213–221 (1994).
54. Pozner-Moulis, S., Cregger, M., Camp, R. L. & Rimm, D. L. Antibody validation by quantitative analysis of protein expression using expression of Met in breast cancer as a model. *Lab. Invest.* **87,** 251–260 (2007).
55. Orlandi, R., Güssow, D. H., Jones, P. T. & Winter, G. Cloning immunoglobulin variable domains for expression by the polymerase chain reaction. *Proc. Natl. Acad. Sci. U. S. A.* **86,** 3833–3837 (1989).
56. Sastry, L., Hay, B. N., Janda, K. D. & Lerner, R. A. Cloning of the immunological repertoire in *Escherichia coli* for generation of monoclonal catalytic antibodies : construction of a heavy chain variable region-specific cDNA library. **86,** 5728–5732 (1989).
57. Larrick, J. W. *et al.* Rapid cloning of rearranged immunoglobulin genes from human hybridoma cells using mixed primers and the polymerase chain reaction. *Biochem. Biophys. Res. Commun.* **160,** 1250–1256 (1989).
58. Sharon, J., Geftert, M. L., Mansert, T. I. M. & Ptashne, M. Site-directed mutagenesis of an invariant amino acid residue at the variable-diversity segments junction of an antibody. **83,** 2628–2631 (1986).

59. Hoogenboom, H. R. & Winter, G. By-passing immunisation. Human antibodies from synthetic repertoires of germline VH gene segments rearranged *in vitro*. *J. Mol. Biol.* **227,** 381–388 (1992).
60. Jones, P. T., Dear, P. H., Foote, J., Neuberger, M. S. & Winter, G. Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature* **321,** 522–525
61. Cabillytt, S. *et al.* Generation of antibody activity from immunoglobulin polypeptide chains produced in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **81,** 3273–3277 (1984).
62. Horwitz, A. H., Chang, C. P., Better, M., Hellstromt, K. E. & Robinson, R. R. Secretion of functional antibody and Fab fragment froom yeast cells. *Proc. Natl. Acad. Sci. U. S. A.* **85,** 8678–8682 (1988).
63. Huston, J. S. *et al.* Protein engineering of antibody binding sites : recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in *Escherichia coli*. **85,** 5879–5883 (1988).
64. Skerra, A. & Plückthun, A. Assembly of a functional immunoglobulin Fv fragment in *Escherichia coli*. *Science* **240,** 1038–1041 (1988).
65. Bird, R. E. *et al.* Single-chain antigen-binding proteins. *Science* **242,** 423–426 (1988).
66. Glockshuber, R., Malia, M., Pfitzinger, I. & Plückthun, A. A comparison of strategies to stabilize immunoglobulin Fv-fragments. *Biochemistry* (1990). doi:10.1021/bi00458a002
67. Ward, E. S., Güssow, D., Griffiths, a D., Jones, P. T. & Winter, G. Binding activities of a repertoire of single immunoglobulin variable domains secreted from *Escherichia coli*. *Nature* **341,** 544–546 (1989).
68. Holliger, P., Prospero, T. & Winter, G. 'Diabodies': small bivalent and bispecific antibody fragments. *Proc. Natl. Acad. Sci. U. S. A.* **90,** 6444–6448 (1993).
69. Iliades, P., Kortt, A. a. & Hudson, P. J. Triabodies: Single chain Fv fragments without a linker form trivalent trimers. *FEBS Lett.* **409,** 437–441 (1997).
70. Huse, W. D. *et al.* Generation of a large combinatorial library of the immunoglobulin repertoire in phage lambda. *Science* **246,** 1275–1281 (1989).
71. Caton, A. J. & Koprowski, H. Influenza virus hemagglutinin-specific antibodies isolated from a combinatorial expression library are closely related to the immune response of the donor. *Proc. Natl. Acad. Sci. U. S. A.* **87,** 6450–6454 (1990).
72. McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* (1990).
73. Clackson, T., Hoogenboom, H. R., Griffiths, a D. & Winter, G. Making antibody fragments using phage display libraries. *Nature* **352,** 624–628 (1991).
74. Smith, G. P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228,** 1315–1317 (1985).
75. Marks, J. D. *et al.* By-passing immunization. Human antibodies from V-gene libraries displayed on phage. *J. Mol. Biol.* **222,** 581–597 (1991).
76. Hoogenboom, H. R. *et al.* Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains. *Nucleic Acids Res.* **19,** 4133–4137 (1991).
77. Garrard, L. J., Yang, M., O'Connell, M. P., Kelley, R. F. & Henner, D. J. Fab assembly and enrichment in a monovalent phage display system. *Biotechnol. (Nature Publ. Company)* **9,** 1373–1377 (1991).

78. McGuinness, B. T. *et al.* Phage diabody repertoires for selection of large numbers of bispecific antibody fragments. *Nat. Biotechnol.* **14,** 1149–1154 (1996).
79. Griffiths, A. D. & Duncan, A. R. Strategies for selection of antibodies by phage display. *Curr. Opin. Biotechnol.* **9,** 102–108 (1998).
80. Vaughan, T. J. *et al.* Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage display library. *Nat. Biotechnol.* **14,** 309–314 (1996).
81. Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nature* (1997).
82. Feldhaus, M. J. *et al.* Flow-cytometric isolation of human antibodies from a nonimmune *Saccharomyces cerevisiae* surface display library. *Nat. Biotechnol.* **21,** 163–70 (2003).
83. Benatuil, L., Perez, J. M., Belk, J. & Hsieh, C. M. An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng. Des. Sel.* **23,** 155–159 (2010).
84. Bowley, D. R., Labrijn, A. F., Zwick, M. B. & Burton, D. R. Antigen selection from an HIV-1 immune antibody library displayed on yeast yields many novel antibodies compared to selection from the same library displayed on phage. *Protein Eng. Des. Sel.* **20,** 81–90 (2007).
85. VanAntwerp, J. J. & Wittrup, K. D. Fine affinity discrimination by yeast surface display and flow cytometry. *Biotechnol. Prog.* **16,** 31–7 (2000).
86. Feldhaus, M. J. & Siegel, R. W. Yeast display of antibody fragments: a discovery and characterization platform. *J. Immunol. Methods* **290,** 69–80 (2004).
87. Boder, E. T., Midelfort, K. S. & Wittrup, K. D. Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 10701–5 (2000).
88. Daugherty, P. S., Chen, G., Olsen, M. J., Iverson, B. L. & Georgiou, G. Antibody affinity maturation using bacterial surface display. *Protein Eng.* **11,** 825–832 (1998).
89. Daugherty, P. S., Olsen, M. J., Iverson, B. L. & Georgiou, G. Development of an optimized expression system for the screening of antibody libraries displayed on the *Escherichia coli* surface. *Protein Eng.* **12,** 613–621 (1999).
90. Gunneriusson, E., Samuelson, P., Uhleń, M., Nygren, P. Å. & Ståhl, S. Surface display of a functional single-chain Fv antibody on staphylococci. *J. Bacteriol.* **178,** 1341–1346 (1996).
91. Du, C., Chan, W. C., McKeithan, T. W. & Nickerson, K. W. Surface display of recombinant proteins on *Bacillus thuringiensis* spores. *Appl. Environ. Microbiol.* **71,** 3337–3341 (2005).
92. Urban, J. H. *et al.* Selection of functional human antibodies from retroviral display libraries. *Nucleic Acids Res.* **33,** 1–8 (2005).
93. Hanes, J. & Plückthun, A. In vitro selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 4937–4942 (1997).
94. Russell, S. J., Hawkins, R. E. & Winter, G. Retroviral vectors displaying functional antibody fragments. *Nucleic Acids Res.* **21,** 1081–1085 (1993).
95. Ho, M., Nagata, S. & Pastan, I. Isolation of anti-CD22 Fv with high affinity by Fv display on human cells. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 9637–9642 (2006).
96. Rode, H. J. *et al.* Cell surface display of a single-chain antibody for attaching polypeptides. *Biotechniques* **21,** 650–658 (1996).

97. Leong, S. S. J. & Chen, W. N. Preparing recombinant single chain antibodies. *Chem. Eng. Sci.* **63,** 1401–1414 (2008).
98. Shusta, E. V, Raines, R. T., Plückthun, A. & Wittrup, K. D. Increasing the secretory capacity of *Saccharomyces cerevisiae* for production of single-chain antibody fragments. *Nat. Biotechnol.* **16,** 773–777 (1998).
99. Wentz, A. E. & Shusta, E. V. A novel high-throughput screen reveals yeast genes that increase secretion of heterologous proteins. *Appl. Environ. Microbiol.* **73,** 1189–1198 (2007).
100. Huang, D. & Shusta, E. V. A yeast platform for the production of single-chain antibody-green fluorescent protein fusions. *Appl. Environ. Microbiol.* **72,** 7748–7759 (2006).
101. Shusta, E. V, Kieke, M. C., Parke, E., Kranz, D. M. & Wittrup, K. D. Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency. *J. Mol. Biol.* **292,** 949–956 (1999).
102. Miller, K. D., Weaver-Feldhaus, J., Gray, S. A., Siegel, R. W. & Feldhaus, M. J. Production, purification, and characterization of human scFv antibodies expressed in *Saccharomyces cerevisiae*, *Pichia pastoris*, and *Escherichia coli*. *Protein Expr. Purif.* **42,** 255–267 (2005).
103. Lesh, F. A. Massive development of amebas in the large intestine. *Am. J. Trop. Med. Hyg.* **24,** 383–392 (1975).
104. World Health Organization. Amoebiasis. *Wkly. Epidemiol. Rec.* **72,** 97–100 (1997).
105. Walsh, J. A. Problems in recognition and diagnosis of amebiasis: estimation of the global magnitude of morbidity and mortality. *Rev. Infect. Dis.* **8,** 228–238 (1986).
106. Gunther, J., Shafir, S., Bristow, B. & Sorvillo, F. Short Report : Amebiasis-related mortality among United States residents, 1990 – 2007. *Am. J. Trop. Med. Hyg.* **85,** 1038–1040 (2011).
107. Gathiram, V. & Jackson, T. F. H. G. A longitudinal study of asymptomatic carriers of pathogenic zymodemes of *Entamoeba histolytica*. *South African Med. J.* **72,** 669–672 (1987).
108. Blessmann, J. *et al.* Longitudinal study of intestinal *Entamoeba histolytica* infections in asymptomatic adult carriers. *J. Clin. Microbiol.* **41,** 4745–4750 (2003).
109. Haque, R. *et al. Entamoeba histolytica* infection in children and protection from subsequent amebiasis. *Infect. Immun.* **74,** 904–909 (2006).
110. Irusen, E. M., Jackson, T. F. & Simjee, A. E. Asymptomatic intestinal colonization by pathogenic *Entamoeba histolytica* in amebic liver abscess: prevalence, response to therapy, and pathogenic potential. *Clin. Infect. Dis.* **14,** 889–893 (1992).
111. Haque, R., Huston, C. D., Hughes, M., Houpt, E. & Petri, W. A. Amebiasis. *N. Engl. J. Med.* **348,** 1565–1573 (2003).
112. Haque, R. *et al.* Innate and acquired resistance to amebiasis in Bangladeshi children. *J. Infect. Dis.* **186,** 547–552 (2002).
113. Salit, I. E., Khairnar, K., Gough, K. & Pillai, D. R. A possible cluster of sexually transmitted *Entamoeba histolytica*: genetic analysis of a highly virulent strain. *Clin. Infect. Dis.* **49,** 346–353 (2009).
114. Stanley, S. L. Amoebiasis. *Lancet* **361,** 1025–1034 (2003).
115. Walderich, B., Burchard, G. D., Knobloch, J. & Müller, L. Development of monoclonal antibodies specifically recognizing the cyst stage of *Entamoeba histolytica*. *Am. J. Trop. Med. Hyg.* **59,** 347–351 (1998).

116. Boggild, A. K., Yohanna, S., Keystone, J. S. & Kain, K. C. Prospective analysis of parasitic infections in Canadian travelers and immigrants. *J. Travel Med.* **13,** 138–144 (2006).
117. Barwick, R. S. *et al.* Outbreak of amebiasis in Tbilisi, Republic of Georgia, 1998. *Am. J. Trop. Med. Hyg.* **67,** 623–631 (2002).
118. Nishise, S. *et al.* Mass infection with *Entamoeba histolytica* in a Japanese institution for individuals with mental retardation: epidemiology and control measures. *Ann. Trop. Med. Parasitol.* **104,** 383–390 (2010).
119. Hung, C.-C., Chang, S.-Y. & Ji, D.-D. *Entamoeba histolytica* infection in men who have sex with men. *Lancet. Infect. Dis.* **12,** 729–36 (2012).
120. Trissl, D. Immunology of *Entamoeba histolytica* in human and animal hosts. *Rev. Infect. Dis.* **4,** 1154–1184
121. Petri, W. A. *et al.* Enteric infections, diarrhea, and their impact on function and development. *J. Clin. Invest.* **118,** 1277–1290 (2008).
122. UNICEF & WHO. *Diarrhea: Why children are still dying and what can be done?* (2009). doi:ISBN 978-92-806-4462-3 (UNICEF)
123. Petri, W. A., Haque, R. & Mann, B. J. The bittersweet interface of parasite and host: lectin-carbohydrate interactions during human invasion by the parasite *Entamoeba histolytica*. *Annu. Rev. Microbiol.* **56,** 39–64 (2002).
124. Mortimer, L. & Chadee, K. The immunopathogenesis of *Entamoeba histolytica*. *Exp. Parasitol.* **126,** 366–380 (2010).
125. Ralston, K. S. & Petri, W. A. Tissue destruction and invasion by *Entamoeba histolytica*. *Trends in Parasitology* **27,** 254–263 (2011).
126. Ralston, K. S. *et al.* Trogocytosis by *Entamoeba histolytica* contributes to cell killing and tissue invasion. *Nature* **508,** 526–30 (2014).
127. Choi, M. H. *et al.* An unusual surface peroxiredoxin protects invasive *Entamoeba histolytica* from oxidant attack. *Mol. Biochem. Parasitol.* **143,** 80–89 (2005).
128. Wilson, I. W., Weedall, G. D. & Hall, N. Host-parasite interactions in *Entamoeba histolytica* and *Entamoeba dispar*: what have we learned from their genomes? *Parasite Immunol.* **34,** 90–99 (2012).
129. Zhang, X. *et al.* Expression of amoebapores is required for full expression of *Entamoeba histolytica* virulence in amebic liver abscess but is not necessary for the induction of inflammation or tissue damage in amebic colitis. *Infect. Immun.* **72,** 678–683 (2004).
130. Haque, R. *et al.* Amebiasis and mucosal IgA antibody against the *Entamoeba histolytica* adherence lectin in Bangladeshi children. *J. Infect. Dis.* **183,** 1787–1793 (2001).
131. Snow, M., Chen, M., Guo, J., Atkinson, J. & Stanley, S. L. Short report: Differences in complement-mediated killing of Entamoeba histolytica between men and women - an explanation for the increased susceptibility of men to invasive amebiasis? *Am. J. Trop. Med. Hyg.* **78,** 922–923 (2008).
132. Duggal, P. *et al.* A mutation in the leptin receptor is associated with *Entamoeba histolytica* infection in children. *J. Clin. Invest.* **121,** 1191–1198 (2011).
133. Guo, X. *et al.* Leptin signaling in intestinal epithelium mediates resistance to enteric infection by *Entamoeba histolytica*. *Mucosal Immunol.* **4,** 294–303 (2011).
134. Galván-Moroyoqui, J. M., del Carmen Domínguez-Robles, M., Franco, E. & Meza, I. The interplay between *Entamoeba* and enteropathogenic bacteria modulates epithelial cell damage. *PLoS Negl. Trop. Dis.* **2,** (2008).

135. Padilla-Vaca, F., Ankri, S., Bracha, R., Koole, L. A. & Mirelman, D. Down regulation of *Entamoeba histolytica* virulence by monoxenic cultivation with *Escherichia coli* O55 is related to a decrease in expression of the light (35-kilodalton) subunit of the Gal/GalNac lectin. *Infect. Immun.* **67,** 2096–2102 (1999).
136. Singh, U. & Ehrenkaufer, G. M. Recent insights into *Entamoeba* development: identification of transcriptional networks associated with stage conversion. *International Journal for Parasitology* **39,** 41–47 (2009).
137. Clark, C. G. & Diamond, L. S. Methods for cultivation of luminal parasitic protists of clinical importance. *Clinical Microbiology Reviews* **15,** 329–341 (2002).
138. Ehrenkaufer, G. M., Haque, R., Hackney, J. A., Eichinger, D. J. & Singh, U. Identification of developmentally regulated genes in *Entamoeba histolytica*: insights into mechanisms of stage conversion in a protozoan parasite. *Cell. Microbiol.* **9,** 1426–1444 (2007).
139. Aguilar-Díaz, H., Díaz-Gallardo, M., Laclette, J. P. & Carrero, J. C. In vitro induction of *Entamoeba histolytica* cyst-like structures from trophozoites. *PLoS Negl. Trop. Dis.* **4,** (2010).
140. Barrón-González, M. P., Villarreal-Treviño, L., Reséndez-Pérez, D., Mata-Cárdenas, B. D. & Morales-Vallarta, M. R. *Entamoeba histolytica*: cyst-like structures *in vitro* induction. *Exp. Parasitol.* **118,** 600–603 (2008).
141. Campos-Góngora, E. *et al.* Mg, Mn, and Co ions enhance the formation of *Entamoeba histolytica* cyst-like structures resistant to sodium dodecyl sulfate. *Arch. Med. Res.* **31,** 162–168 (2000).
142. Eichinger, D. A role for a galactose lectin and its ligands during encystment of *Entamoeba*. *J. Eukaryot. Microbiol.* **48,** 17–21 (2001).
143. Chatterjee, A. *et al.* Evidence for a 'wattle and daub' model of the cyst wall of *Entamoeba*. *PLoS Pathog.* **5,** (2009).
144. Samuelson, J. & Robbins, P. W. A simple fibril and lectin model for cyst walls of *Entamoeba* and perhaps *Giardia*. *Trends Parasitol.* **27,** 17–22 (2011).
145. Frisardi, M. *et al.* The most abundant glycoprotein of amebic cyst walls (Jacob) is a lectin with five Cys-rich, chitin-binding domains. *Infect. Immun.* **68,** 4217–4224 (2000).
146. Samuelson, J., Bushkin, G. G., Chatterjee, A. & Robbins, P. W. Strategies to discover the structural components of cyst and oocyst walls. *Eukaryotic Cell* **12,** 1578–1587 (2013).
147. Blessmann, J. & Tannich, E. Treatment of asymptomatic intestinal *Entamoeba histolytica* infection. *The New England journal of medicine* **347,** (2002).
148. Debnath, A. *et al.* A high-throughput drug screen for *Entamoeba histolytica* identifies a new lead and target. *Nature Medicine* **18,** 956–960 (2012).
149. Quach, J., St-Pierre, J. & Chadee, K. The future for vaccine development against *Entamoeba histolytica*. *Human Vaccines and Immunotherapeutics* **10,** 1514–1521 (2014).
150. Fotedar, R. *et al.* Laboratory diagnostic techniques for *Entamoeba* species. *Clin. Microbiol. Rev.* **20,** 511–532 (2007).
151. Stark, D. *et al.* Evaluation of multiplex tandem real-time PCR for detection of *Cryptosporidium* spp., *Dientamoeba fragilis*, *Entamoeba histolytica*, and *Giardia intestinalis* in clinical stool samples. *J. Clin. Microbiol.* **49,** 257–262 (2011).
152. Haque, R., Neville, L. M., Hahn, P. & Petri, W. A. Rapid diagnosis of *Entamoeba* infection by using *Entamoeba* and *Entamoeba histolytica* stool antigen detection kits. *J. Clin. Microbiol.* **33,** 2558–2561 (1995).

153. Blessmann, J. *et al.* Real-time PCR for detection and differentiation of *Entamoeba histolytica* and *Entamoeba dispar* in fecal samples. *J. Clin. Microbiol.* **40,** 4413–4417 (2002).
154. Diamond, L. S. & Clark, C. G. A redescription of *Entamoeba histolytica* Schaudinn, 1903 (Emended Walker, 1911) separating it from *Entamoeba dispar* Brumpt, 1925. *J. Eukaryot. Microbiol.* **40,** 340–344 (1993).
155. Clark, C. G. & Diamond, L. S. The Laredo strain and other '*Entamoeba histolytica*-like' amoebae are *Entamoeba moshkovskii*. *Mol. Biochem. Parasitol.* **46,** 11–18 (1991).
156. Jackson, T. F., Gathiram, V. & Simjee, A. E. Seroepidemiological study of antibody responses to the zymodemes of *Entamoeba histolytica*. *Lancet* **1,** 716–719 (1985).
157. Knobloch, J. & Mannweiler, E. Development and persistence of antibodies to *Entamoeba histolytica* in patients with amebic liver abscess. Analysis of 216 cases. *Am. J. Trop. Med. Hyg.* **32,** 727–732 (1983).
158. Gatti, S. *et al.* Amebic infections due to the *Entamoeba histolytica-Entamoeba dispar* complex: a study of the incidence in a remote rural area of Ecuador. *Am. J. Trop. Med. Hyg.* **67,** 123–127 (2002).
159. Mirelman, D., Nuchamowitz, Y. & Stolarsky, T. Comparison of use of enzyme-linked immunosorbent assay-based kits and PCR amplification of rRNA genes for simultaneous detection of *Entamoeba histolytica* and *E. dispar*. *J. Clin. Microbiol.* **35,** 2405–2407 (1997).
160. Stark, D. *et al.* Comparison of stool antigen detection kits to PCR for diagnosis of amebiasis. *J. Clin. Microbiol.* **46,** 1678–1681 (2008).
161. Roy, S. *et al.* Real-time-PCR assay for diagnosis of *Entamoeba histolytica* infection. *J. Clin. Microbiol.* **43,** 2168–2172 (2005).
162. Ptolemy, A. S. & Rifai, N. What is a biomarker? Research investments and lack of clinical integration necessitate a review of biomarker terminology and validation schema. *Scand. J. Clin. Lab. Invest. Suppl.* **242,** 6–14 (2010).
163. Atkinson A.J., J. *et al.* Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* **69,** 89–95 (2001).
164. Strimbu, K. & Tavel, J. A. What are biomarkers? *Curr. Opin. HIV AIDS* **5,** 463–466 (2010).
165. Tachibana, H. *et al.* Identification of a pathogenic isolate-specific 30,000-M(r) antigen of *Entamoeba histolytica* by using a monoclonal antibody. *Infect. Immun.* **58,** 955–960 (1990).
166. Tachibana, H. *et al.* Differences in genomic DNA sequences between pathogenic and nonpathogenic isolates of *Entamoeba histolytica* identified by polymerase chain reaction. *J. Clin. Microbiol.* **29,** 2234–2239 (1991).
167. Torian, B. E., Flores, B. M., Stroeher, V. L., Hagen, F. S. & Stamm, W. E. cDNA sequence analysis of a 29-kDa cysteine-rich surface antigen of pathogenic Entamoeba histolytica. *Proc. Natl. Acad. Sci. U. S. A.* **87,** 6358–6362 (1990).
168. Reed, S. L. *et al.* Molecular and cellular characterization of the 29-kilodalton peripheral membrane protein of *Entamoeba histolytica*: differentiation between pathogenic and nonpathogenic isolates. *Infect. Immun.* **60,** 542–549 (1991).
169. Tachibana, H., Kobayashi, S., Cheng, X. J. & Hiwatashi, E. Differentiation of *Entamoeba histolytica* from *E. dispar* facilitated by monoclonal antibodies against a 150-kDa surface antigen. *Parasitol. Res.* **83,** 435–439 (1997).

170. Cheng, X. J., Tsukamoto, H., Kaneda, Y. & Tachibana, H. Identification of the 150-kDa surface antigen of *Entamoeba histolytica* as a galactose- and N-acetyl-D-galactosamine-inhibitable lectin. *Parasitol. Res.* **84,** 632–639 (1998).
171. Petri, W. A. *et al.* Pathogenic and nonpathogenic strains of *Entamoeba histolytica* can be differentiated by monoclonal antibodies to the galactose-specific adherence lectin. *Infect. Immun.* **58,** 1802–1806 (1990).
172. Moody, S., Becker, S., Nuchamowitz, Y. & Mirelman, D. Virulent and avirulent *Entamoeba histolytica* and *E. dispar* differ in their cell surface phosphorylated glycolipids. *Parasitology* **114,** 95–104 (1997).
173. Bracha, R., Nuchamowitz, Y. & Mirelman, D. Molecular cloning of a 30-kilodalton lysine-rich surface antigen from a nonpathogenic *Entamoeba histolytica* strain and its expression in a pathogenic strain. *Infect. Immun.* **63,** 917–925 (1995).
174. Strachan, W. D., Chiodini, P. L., Spice, W. M., Moody, A. H. & Ackers, J. P. Immunological differentiation of pathogenic and non-pathogenic isolates of *Entamoeba histolytica*. *Lancet* **1,** 561–563 (1988).
175. Tanyuksel, M. & Petri, W. A. Laboratory diagnosis of amebiasis. *Clin. Microbiol. Rev.* **16,** 713–729 (2003).
176. Pillai, D. R. & Kain, K. C. Immunochromatographic strip-based detection of *Entamoeba histolytica-E. dispar* and *Giardia lamblia* coproantigen. *J. Clin. Microbiol.* **37,** 3017–3019 (1999).
177. Garcia, L. S., Shimizu, R. Y. & Bernard, C. N. Detection of *Giardia lamblia*, *Entamoeba histolytica*/*Entamoeba dispar*, and *Cryptosporidium parvum* antigens in human fecal specimens using the triage parasite panel enzyme immunoassay. *J. Clin. Microbiol.* **38,** 3337–3340 (2000).
178. Stanley, S. L., Becker, A., Kunz-Jenkins, C., Foster, L. & Li, E. Cloning and expression of a membrane antigen of *Entamoeba histolytica* possessing multiple tandem repeats. *Proc. Natl. Acad. Sci. U. S. A.* **87,** 4976–4980 (1990).
179. Pillai, D. R. *et al. Entamoeba histolytica* and *Entamoeba dispar*: epidemiology and comparison of diagnostic methods in a setting of nonendemicity. *Clin. Infect. Dis.* **29,** 1315–1318 (1999).
180. Haque, R. *et al.* Diagnosis of pathogenic *Entamoeba histolytica* infection using a stool ELISA based on monoclonal antibodies to the galactose-specific adhesin. *J. Infect. Dis.* **167,** 247–249 (1993).
181. Haque, R. *et al.* Diagnosis of amebic liver abscess and intestinal infection with the TechLab *Entamoeba histolytica* II antigen detection and antibody tests. *J. Clin. Microbiol.* **38,** 3235–3239 (2000).
182. Mann, B. J. *et al.* Neutralizing monoclonal antibody epitopes of the *Entamoeba histolytica* galactose adhesin map to the cysteine-rich extracellular domain of the 170-kilodalton subunit. *Infect. Immun.* **61,** 1772–8 (1993).
183. Dodson, J. M. *et al.* Comparison of adherence, cytotoxicity, and Gal/GalNAc lectin gene structure in *Entamoeba histolytica* and *Entamoeba dispar*. *Parasitology International* **46,** 225–235 (1997).
184. Beck, D. L. *et al. Entamoeba histolytica*: sequence conservation of the Gal/GalNAc lectin from clinical isolates. *Exp. Parasitol.* **101,** 157–163 (2002).

185. Prim, N. *et al.* Risk of underdiagnosing amebic dysentery due to false-negative *Entamoeba histolytica* antigen detection. *Diagn. Microbiol. Infect. Dis.* **73,** 372–373 (2012).
186. Ali, I. K. M. *et al.* Proteomic analysis of the cyst stage of *Entamoeba histolytica*. *PLoS Negl. Trop. Dis.* **6,** (2012).
187. Guild, K. *et al.* Wheat germ cell-free expression system as a pathway to improve protein yield and solubility for the SSGCID pipeline. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **67,** 1027–1031 (2011).
188. Biller, L. *et al.* The cell surface proteome of *Entamoeba histolytica*. *Mol. Cell. Proteomics* **13,** 132–44 (2014).
189. Hon, C. C. *et al.* Quantification of stochastic noise of splicing and polyadenylation in *Entamoeba histolytica*. *Nucleic Acids Res.* **41,** 1936–1952 (2013).
190. Debnath, A. *et al.* Hsp90 inhibitors as new leads to target parasitic diarrheal diseases. *Antimicrob. Agents Chemother.* **58,** 4138–4144 (2014).
191. Gonzalez, E. *et al.* Calreticulin-like molecule in trophozoites of *Entamoeba histolytica* HM1:IMSS (SWISSPROT: ACCESSION P83003). *Am. J. Trop. Med. Hyg.* **67,** 636–639 (2002).
192. Vaithilingam, A., Teixeira, J. E., Miller, P. J., Heron, B. T. & Huston, C. D. *Entamoeba histolytica* cell surface calreticulin binds human C1q and functions in amebic phagocytosis of host cells. *Infect. Immun.* **80,** 2008–2018 (2012).
193. Ghosh, S. K. *et al.* The Jacob2 lectin of the *Entamoeba histolytica* cyst wall binds chitin and is polymorphic. *PLoS Negl. Trop. Dis.* **4,** 3–9 (2010).
194. Lubec, G. & Afjehi-Sadat, L. Limitations and pitfalls in protein identifications by mass spectrometry. *Chemical Reviews* **107,** 3568–3584 (2007).
195. Hall, T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41,** 95–98 (1999).
196. Bendtsen, J. D., Nielsen, H., Von Heijne, G. & Brunak, S. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340,** 783–795 (2004).
197. Gray, S. A. *et al.* Toward low-cost affinity reagents: lyophilized yeast-scFv probes specific for pathogen antigens. *PLoS One* **7,** (2012).
198. Myler, P. J. *et al.* The Seattle Structural Genomics Center for Infectious Disease (SSGCID). *Infect. Disord. Drug Targets* **9,** 493–506 (2009).
199. Hewitt, S. N. *et al.* Expression of proteins in *Escherichia coli* as fusions with maltose-binding protein to rescue non-expressed targets in a high-throughput protein-expression and purification pipeline. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **67,** 1006–1009 (2011).
200. Beckett, D., Kovaleva, E. & Schatz, P. J. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci.* **8,** 921–929 (1999).
201. Choi, R. *et al.* Immobilized metal-affinity chromatography protein-recovery screening is predictive of crystallographic structure success. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **67,** 998–1005 (2011).
202. Gai, S. A. & Wittrup, K. D. Yeast surface display for protein engineering and characterization. *Curr. Opin. Struct. Biol.* **17,** 467–73 (2007).
203. Weaver-Feldhaus, J. M. *et al.* Yeast mating for combinatorial Fab library generation and surface display. *FEBS Lett.* **564,** 24–34 (2004).

204. Walker, L. M., Bowley, D. R. & Burton, D. R. Efficient recovery of high-affinity antibodies from a single-chain Fab yeast display library. *J. Mol. Biol.* **389,** 365–375 (2009).
205. Lee, H.-W. *et al.* Construction and characterization of a pseudo-immune human antibody library using yeast surface display. *Biochem. Biophys. Res. Commun.* **346,** 896–903 (2006).
206. Xin, X. W. & Shusta, E. V. The use of scFv-displaying yeast in mammalian cell surface selections. *J. Immunol. Methods* **304,** 30–42 (2005).
207. Yeung, Y. A. & Wittrup, K. D. Quantitative screening of yeast surface-displayed polypeptide libraries by magnetic bead capture. *Biotechnol. Prog.* **18,** 212–220 (2002).
208. Siegel, R. W., Coleman, J. R., Miller, K. D. & Feldhaus, M. J. High efficiency recovery and epitope-specific sorting of an scFv yeast display library. *J. Immunol. Methods* **286,** 141–53 (2004).
209. Boder, E. T. & Wittrup, K. D. Optimal screening of surface-displayed polypeptide libraries. *Biotechnol. Prog.* **14,** 55–62 (1998).
210. Cho, Y. K., Chen, I., Wei, X., Li, L. & Shusta, E. V. A yeast display immunoprecipitation method for efficient isolation and characterization of antigens. *J. Immunol. Methods* **341,** 117–126 (2009).
211. Orr, B. a., Carr, L. M., Wittrup, K. D., Roy, E. J. & Kranz, D. M. Rapid method for measuring scFv thermal stability by yeast surface display. *Biotechnol. Prog.* **19,** 631–638 (2003).
212. Gray, S. A. *et al.* Flow cytometry-based methods for assessing soluble scFv activities and detecting antigens in solution. *Biotechnol. Bioeng.* **105,** 973–981 (2010).
213. Kahn, M. *et al.* Improved methodology to preserve yeast-bound single chain fragment variable antibody affinity reagents. Submitted.
214. Luo, X. & Davis, J. J. Electrical biosensors and the label free detection of protein disease biomarkers. *Chem. Soc. Rev.* **42,** 5944–62 (2013).
215. Grewal, Y. S. *et al.* Label-free electrochemical detection of an *Entamoeba histolytica* antigen using cell-free yeast-scFv probes. *Chem. Commun. (Camb).* **49,** 1551–3 (2013).
216. Grewal, Y. S., Shiddiky, M. J. a, Spadafora, L. J., Cangelosi, G. A. & Trau, M. Nano-yeast-scFv probes on screen-printed gold electrodes for detection of *Entamoeba histolytica* antigens in a biological matrix. *Biosens. Bioelectron.* **55,** 417–422 (2014).
217. Wang, Y. *et al.* Duplex microfluidic SERS detection of pathogen antigens with nanoyeast single-chain variable fragments. *Anal. Chem.* **86,** 9930–8 (2014).
218. Kneipp, J., Kneipp, H. & Kneipp, K. SERS--a single-molecule and nanoscale tool for bioanalytics. *Chem. Soc. Rev.* **37,** 1052–1060 (2008).
219. Venkatesh, A. G. *et al.* Yeast dual-affinity biobricks: Progress towards renewable whole-cell biosensors. *Biosens. Bioelectron.* **70,** 462–468 (2015).
220. Tachibana, H. *et al.* Preparation of recombinant human monoclonal antibody Fab fragments specific for *Entamoeba histolytica*. *Clin. Diagn. Lab. Immunol.* **6,** 383–387 (1999).
221. Tachibana, H. *et al.* Bacterial expression of a neutralizing mouse monoclonal antibody Fab fragment to a 150-kilodalton surface antigen of *Entamoeba histolytica*. *Am. J. Trop. Med. Hyg.* **60,** 35–40 (1999).

222. Tachibana, H. *et al.* VH3 gene usage in neutralizing human antibodies specific for the *Entamoeba histolytica* Gal/GalNAc lectin heavy subunit. *Infect. Immun.* **71,** 4313–4319 (2003).
223. Tachibana, H. *et al.* Bacterial expression of a human monoclonal antibody-alkaline phosphatase conjugate specific for *Entamoeba histolytica*. *Clin. Diagn. Lab. Immunol.* **11,** 216–218 (2004).
224. Houpt, E. *et al.* Prevention of intestinal amebiasis by vaccination with the *Entamoeba histolytica* Gal/GalNac lectin. *Vaccine* **22,** 611–617 (2004).
225. Mondal, D., Petri, W. A., Sack, R. B., Kirkpatrick, B. D. & Haque, R. *Entamoeba histolytica*-associated diarrheal illness is negatively associated with the growth of preschool children: evidence from a prospective study. *Trans. R. Soc. Trop. Med. Hyg.* **100,** 1032–1038 (2006).
226. Mondal, D. *et al.* Contribution of enteric infection, altered intestinal barrier function, and maternal malnutrition to infant malnutrition in Bangladesh. *Clin. Infect. Dis.* **54,** 185–92 (2012).
227. Petri, W. A. *et al. Entamoeba bangladeshi* nov. sp., Bangladesh. *Emerg. Infect. Dis.* **18,** 1543–1545 (2012).
228. Van Dellen, K. L. *et al.* Unique posttranslational modifications of chitin-binding lectins of *Entamoeba invadens* cyst walls. *Eukaryot. Cell* **5,** 836–848 (2006).
229. Ghosh, S. *et al.* Molecular epidemiology of *Entamoeba* spp.: evidence of a bottleneck (demographic sweep) and transcontinental spread of diploid parasites. *J. Clin. Microbiol.* **38,** 3815–3821 (2000).
230. Ayeh-Kumi, P. F. *et al. Entamoeba histolytica*: genetic diversity of clinical isolates from Bangladesh as demonstrated by polymorphisms in the serine-rich gene. *Exp. Parasitol.* **99,** 80–88 (2001).
231. Mondal, D., Haque, R., Sack, R. B., Kirkpatrick, B. D. & Petri, W. A. Attribution of malnutrition to cause-specific diarrheal illness: evidence from a prospective study of preschool children in Mirpur, Dhaka, Bangladesh. *Am. J. Trop. Med. Hyg.* **80,** 824–826 (2009).
232. Robinson, G. L. Laboratory cultivation of some human parasitic amoebae. *J. Gen. Microbiol.* **53,** 69–79 (1968).
233. Diamond, L. S. in *In Vitro Methods for Parasite Cultivation* (eds. Taylor, A. & Baker, J.) 1–28 (Academic Press, 1987).
234. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9,** 671–675 (2012).
235. Grewal, Y. S., Shiddiky, M. J. A., Spadafora, L. J., Cangelosi, G. A. & Trau, M. Structural characterization of nanoyeast single-chain fragment variable affinity reagents. *J. Phys. Chem. C* **119,** 12674–12680 (2015).
236. Cheng, X. J., Ihara, S., Takekoshi, M. & Tachibana, H. *Entamoeba histolytica*: bacterial expression of a human monoclonal antibody which inhibits *in vitro* adherence of trophozoites. *Exp. Parasitol.* **96,** 52–56 (2000).
237. Konthur, Z., Hust, M. & Dübel, S. Perspectives for systematic in vitro antibody generation. in *Gene* **364,** 19–29 (2005).