JCI insight

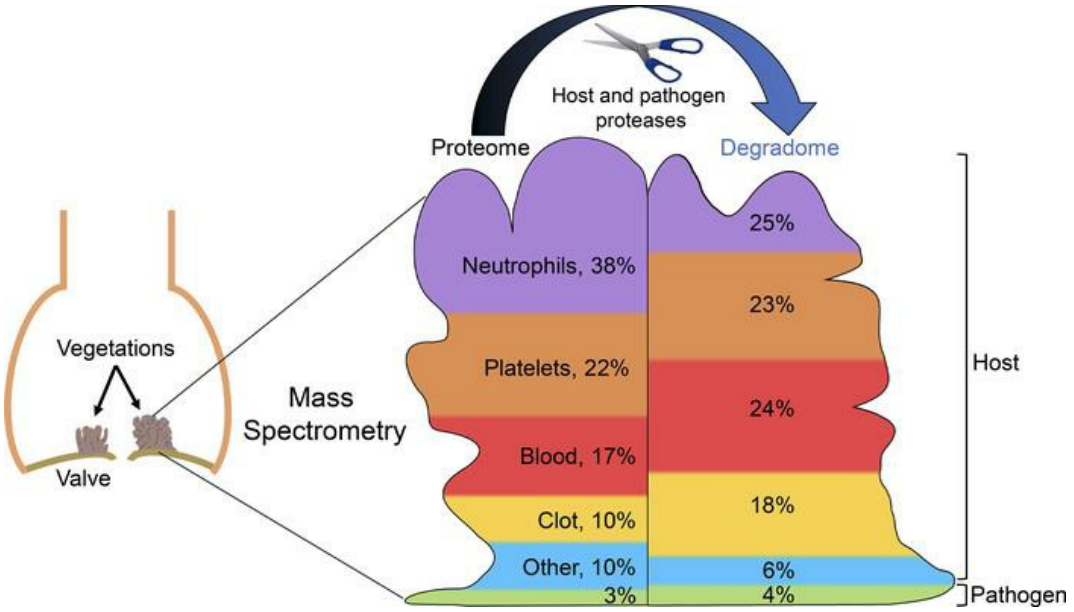
Proteomics identifies a convergent innate response to infective endocarditis and extensive proteolysis in vegetation components

Daniel R. Martin, … , Belinda B. Willard, Suneel S. Apte

JCI Insight. 2020;5(14):e135317. https://doi.org/10.1172/jci.insight.135317.

Research Article | Cardiology | Infectious disease

Graphical abstract



Find the latest version:

https://jci.me/135317/pdf

JCI insight

RESEARCH ARTICLE

Proteomics identifies a convergent innate response to infective endocarditis and extensive proteolysis in vegetation components

Daniel R. Martin,[1] James C. Witten,[2] Carmela D. Tan,[3] E. Rene Rodriguez,[3] Eugene H. Blackstone,[2] Gosta B. Pettersson,[2] Deborah E. Seifert,[1] Belinda B. Willard,[4] and Suneel S. Apte[1]

[1]Department of Biomedical Engineering, Lerner Research Institute, [2]Department of Thoracic and Cardiovascular Surgery, Miller Family Heart and Vascular Institute, [3]Department of Pathology, Robert J. Tomsich Pathology & Laboratory Medicine Institute, and [4]Proteomics and Metabolomics Core, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio, USA.

Infective endocarditis is a life-threatening infection of heart valves and adjacent structures characterized by vegetations on valves and other endocardial surfaces, with tissue destruction and risk of embolization. We used high-resolution mass spectrometry to define the proteome of staphylococcal and non-staphylococcal vegetations and Terminal Amine Isotopic Labeling of Substrates (TAILS) to define their proteolytic landscapes. These approaches identified over 2000 human proteins in staphylococcal and non-staphylococcal vegetations. Individual vegetation proteomes demonstrated comparable profiles of quantitatively major constituents that overlapped with serum, platelet, and neutrophil proteomes. Staphylococcal vegetation proteomes resembled one another more than the proteomes of non-staphylococcal vegetations. TAILS demonstrated extensive proteolysis within vegetations, with numerous previously undescribed cleavages. Several proteases and pathogen-specific proteins, including virulence factors, were identified in most vegetations. Proteolytic peptides in fibronectin and complement C3 were identified as potential infective endocarditis biomarkers. Overlap of staphylococcal and non-staphylococcal vegetation proteomes suggests a convergent thrombotic and immune response to endocardial infection by diverse pathogens. However, the differences between staphylococcal and non-staphylococcal vegetations and internal variance within the non-staphylococcal group indicate that additional pathogen- or patient-specific effects exist. Pervasive proteolysis of vegetation components may arise from vegetation-intrinsic proteases and destabilize vegetations, contributing to embolism.

Introduction

Infective endocarditis is a life-threatening infection of heart valves and other endocardial surfaces characterized by wart-like growths on the valve surface — called vegetations — that may cause extensive valvular destruction. Infective endocarditis can be complicated by formation of abscesses and fistulas affecting the surrounding cardiac structures. Fragments of vegetations may also embolize systemically, causing infarction and/or dissemination of infection. Although the epidemiology of infective endocarditis varies geographically, its incidence is rising globally, owing to an increasing prevalence of risk factors (1–3). Its prognosis depends greatly on comorbidities, the valves affected, the infecting pathogen, and necessity of surgery. Nearly 50% of patients require surgery for complications, and surgery for infective endocarditis carries the greatest risk of any valve surgery (4). Contemporary in-hospital mortality ranges from 15% to 20%, with 1-year mortality at nearly 40% (5–8).

Surprisingly, systems-level analyses of infective endocarditis vegetations are unavailable, although their pathogenesis is well described (9). In 1885, William Osler outlined a pathogenic sequence of infective endocarditis that was based on clinical observation, gross pathology, and microscopic analysis (10), which has been expanded on subsequently and is thought to underlie vegetations regardless of the infecting organism. Damage to valve endothelium, typically occurring in the setting of structural heart disease, turbulent flow, or intravascular devices, exposes collagen and other prothrombogenic extracellular matrix (ECM)

Conflict of interest: The authors have declared that no conflict of interest exists.

Copyright: © 2020, American Society for Clinical Investigation.

Submitted: November 27, 2019
Accepted: June 10, 2020
Published: July 23, 2020.

Reference information: *JCI Insight*. 2020;5(14):e135317. https://doi.org/10.1172/jci.insight.135317.

components of subendothelial layers of native valves, favoring platelet binding. A platelet-rich nidus and serum proteins, including clotting factors, are thought to accumulate in a nonbacterial thrombotic endocarditis (11). This lesion, the original endothelial injury, or prosthetic valve surfaces provide an adhesion substrate for virulence factors of typical infective endocarditis pathogens, such as *Staphylococcus aureus* (11). The vegetations are thought to grow by interactions between the pathogen and innate immune mechanisms that result in accretions of the pathogen, platelets, other cells, clotting factors, and other blood proteins (12).

Elucidation of vegetation proteomes may offer new insights into infective endocarditis pathogenesis, the host response, and infecting microorganisms and identify prospective biomarkers. A previous proteomic analysis of a formalin-fixed, paraffin-embedded vegetation obtained from a rare *Mycobacterium tuberculosis* infective endocarditis case identified only 85 proteins, including 3 of mycobacterial origin (13). Analysis of vegetations by high-resolution mass spectrometry methods has not been undertaken but could provide a systems biology view of their composition, enable comparison of vegetations arising from different microorganisms, and facilitate a complete understanding of vegetation biogenesis. We hypothesized that the protein composition of infective endocarditis vegetations arising from different pathogens may be similar, with observed differences reflecting pathogen-specific responses and individual clinical scenarios. To address this, the proteomes of a diverse cohort of human infective endocarditis vegetations were determined using high-resolution liquid chromatography tandem mass spectrometry (LS-MS/MS) and analyzed with reference to the underlying pathogens, clinical manifestations, and histopathological features. By analyzing vegetations resulting from infection by a typical infective endocarditis pathogen (*S. aureus*, cases designated by SA) and atypical pathogens (non-staphylococcal, cases designated by NSA), this approach unequivocally demonstrated that vegetation composition is highly complex. Because we identified several proteases in the vegetations, an N-terminomics method, Terminal Amine Isotopic Labeling of Substrates (TAILS) (14), was applied to define proteolytic modification of vegetation components, i.e., their degradomes. The respective approaches are relevant to vegetation formation and embolization because TAILS suggested pervasive proteolysis of the vegetation components.

Results

Histological diversity of staphylococcal and non-staphylococcal vegetations. The histopathology summary of each case from which vegetations were analyzed (Table 1) is presented in Supplemental Table 1; supplemental material available online with this article; https://doi.org/10.1172/jci.insight.135317DS1. Vegetation histology revealed acute inflammation, including a neutrophil infiltrate and focal abscesses in all but one (NSA4) vegetation (Figure 1A). Granulation tissue was seen in some native valve vegetations (Supplemental Table 1). Movat pentachrome stain highlighted the valve tissue, when present, staining collagen yellow and elastic fibers black, whereas vegetations lacked specific staining for collagen, elastin, and proteoglycan, indicating their absence (Figure 1A). In contrast, PTAH stain demonstrated abundant fibrin arranged in diverse distributions in the vegetations (Figure 1A). Some vegetations, such as SA2, demonstrated extensive neutrophil infiltrate embedded in a fibrin-enriched matrix, whereas NSA4 had little cell infiltrate, comprising instead large microbial colonies encapsulated in a fibrin network (Figure 1A). Two vegetations (SA2 and NSA4) contained Gram-positive microbes (Supplemental Figure 1A). Neutrophils undergo NETosis in extreme stress situations and release neutrophil extracellular traps (NETs). Neutrophils and NETs were detected in vegetations by CD66 and citrullinated histone H3 immunostaining, respectively (Supplemental Figure 1B).

Proteome of infective endocarditis vegetations. The vegetation proteomes were determined by a shotgun proteomics approach with supplementation of the protein identifications using TAILS. Data for the shotgun and shotgun plus TAILS analyses are presented separately because of an inherently unique aspect of TAILS; i.e., single high-confidence peptides were sought in TAILS database searches as indicators of a proteolytic event, whereas identification of 2 high-confidence peptides was a requirement for identification of proteins in the shotgun approach. Shotgun analysis of the staphylococcal group without fractionation of peptides before LC-MS/MS identified 1073 high-confidence human proteins from 33,410 peptide spectrum matches (PSMs) and 5999 peptide groups (Figure 1B, Supplemental Table 2). Of the collective 1073 proteins, 605 (56%) in the staphylococcal group were present in vegetations from all 3 cases (Figure 1B). Shotgun analysis of the non-staphylococcal vegetations without fractionation of peptides before LC-MS/MS identified 1093 high-confidence human proteins from 40,707 PSMs and 6946 peptide groups (Figure 1B, Supplemental Table 2). Only 216 of the collective 1093 non-staphylococcal proteins (20%) were detected in all the non-staphylococcal vegetations (Figure 1B), and the average overlap in any combination of 3

Table 1. Clinical characteristics of the study population

Label	Age	Sex	Microbial infection	Antibiotic treatment (primary antibiotic bolded)	Preoperative antibiotic treatment (days)	Location of vegetation	Valve type	Other relevant information
SA1	46	M	*Staphylococcus aureus* (methicillin resistant)	**Vancomycin**	12	Base of mitral valve	Native	Injection drug user (cocaine). Symptomatic for 20 days.
SA2	32	F	*Staphylococcus aureus* (methicillin resistant)	Daptomycin, **ceftaroline**	10	Tricuspid valve leaflet	Native	Injection drug user (methamphetamine, opioids), third surgery for infective endocarditis. Symptomatic for 17 days.
SA3	70	F	*Staphylococcus aureus* (methicillin resistant)	**Vancomycin**	42	Vegetations surrounding fistula in right atrium	Not on valve	Complex invasive endocarditis including large fistula from aortic root to right atrium. Symptomatic for 42 days.
NSA1	51	F	*Streptococcus dysgalactiae*	Vancomycin, **ceftriaxone**	30	Aortic valve leaflet	Native	Morbid obesity (BMI > 50), foot osteomyelitis. Symptomatic for 30 days.
NSA2	58	M	*Streptococcus bovis*	Azithromycin, gentamicin, **ceftriaxone**	14	Aortic valve leaflet	Native	Symptomatic for 15 days.
NSA3	70	F	*Candida parapsilosis*	**Micafungin**	13	Aortic valve leaflet	Bioprosthetic	History of *Candida* and enterococcal endocarditis treated with AVR in 2011. Symptomatic for 13 days.
NSA4	78	M	*Cutibacterium acnes* (formerly *Propionibacterium acnes*)	Vancomycin, **ceftriaxone**	5	Aortic valve leaflet	Bioprosthetic	Prolonged subacute symptoms (fatigue/malaise) for 265 days.
NSA5[A]	53	M	*Streptococcus parasanguinis*	Vancomycin, **ceftriaxone**	2	Aortic valve leaflet (A), tricuspid valve leaflet (T)	Native	Bicuspid aortic valve. Symptomatic for 21 days.

Antibiotics are in the order of when they were first administered. [A]Two vegetations from NSA5, from the aortic valve and the tricuspid valve; when considered separately, these vegetations are specified as either aortic (A) or tricuspid (T). SA, *Staphylococcus aureus*; NSA, non-staphylococcal; BMI, body mass index; AVR, aortic valve replacement.

non-staphylococcal cases was 21%. In contrast, 2 vegetations from anatomically distinct sites in the same patient (NSA5) had a proteome overlap of 67% (correlation coefficient of 0.4569, Supplemental Figure 2, A and B). A summary comparison of proteomes of all vegetations is presented in Supplemental Figure 2C. The mass spectrometry data are provided in the Supplemental Data and were deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRoteomics IDEntifications Database partner repository with the data set identifiers PXD016415 for the nonfractionated data and PXD019129 for the fractionated data.

Of the total 1181 proteins in staphylococcal and non-staphylococcal shotgun proteomes, 966 (82%) were found in at least 1 sample from each group (Figure 1C). Plotting the average normalized total ion abundance (from all vegetations) against the number of contributing proteins demonstrated that the 15 most abundant proteins occupied a disproportionately large space (57%) in the vegetation proteomes (Figure 1D). These major components were similar between the staphylococcal and non-staphylococcal groups and included fibrinogen A, vimentin, hemoglobin B, fibrinogen B, and β-actin, although their rank order differed between the groups (Figure 1E). A similar distribution was seen in each individual sample with the top-ranking proteins being mostly identical (Supplemental Figure 3). Conversely, the 584 least abundant proteins contributed to only 1% of the vegetation proteome (Figure 1D).

Protein list analysis using the PANTHER Classification System showed enrichment of similar components and processes in staphylococcal and non-staphylococcal vegetations (Figure 2A and Supplemental Figure 4, A–C). Because vegetations are thought to represent an amalgam of blood proteins, platelet aggregates,

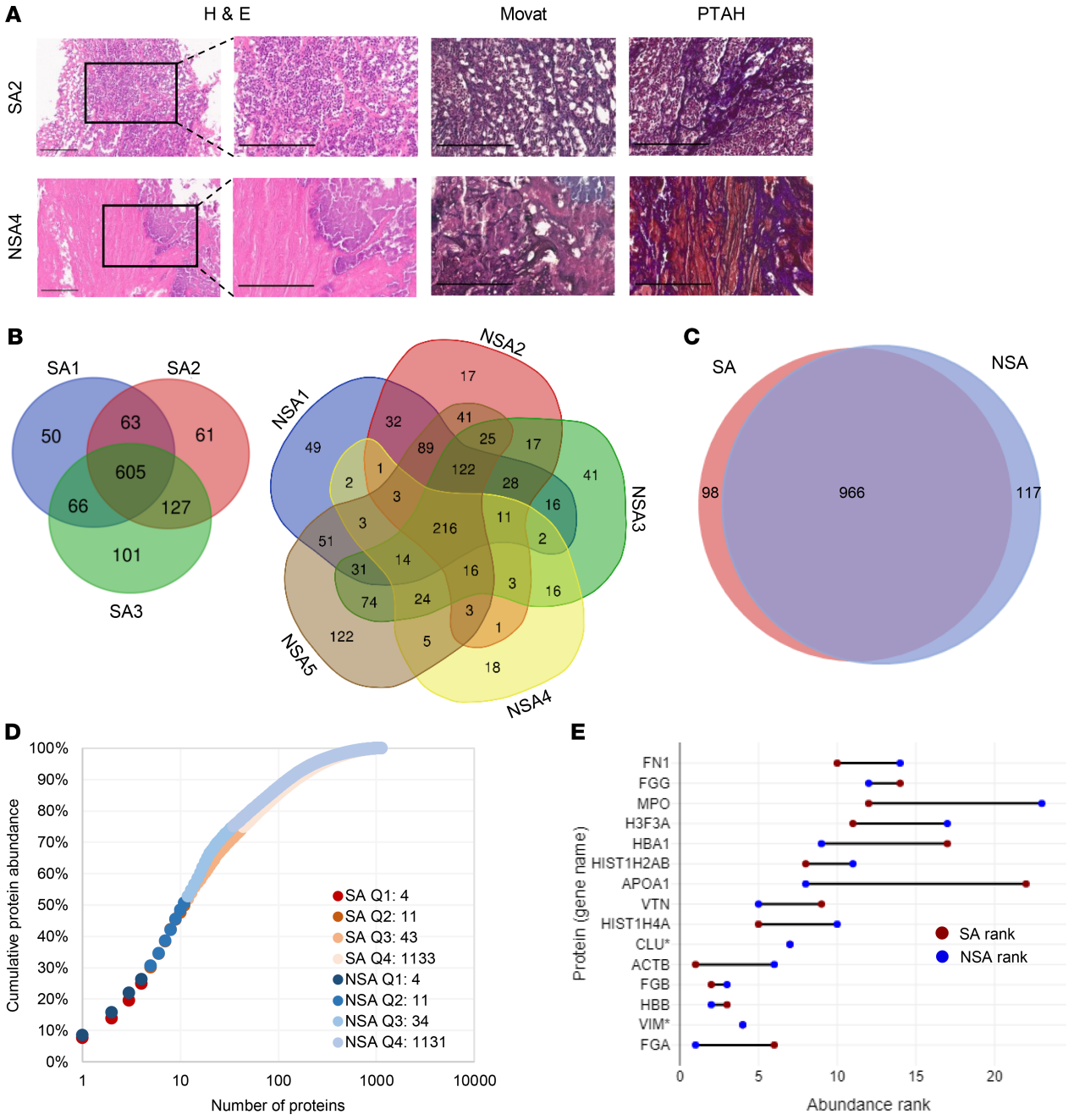


Figure 1. Vegetation histology and shotgun proteome overview. (**A**) Vegetation staining with hematoxylin and eosin (H&E), Movat pentachrome (Movat), and phosphotungstic acid-hematoxylin (PTAH) illustrates contrasting levels of inflammation: numerous inflammatory cells in *S. aureus* 2 (SA2) and few in non–*S. aureus* 4 (NSA4) and abundant fibrin (PTAH stain, shown in purple). Movat staining disclosed no elastic fibers, collagen, or proteoglycans. The boxes in the left-hand panel indicate regions shown at higher magnification in the panels on their right. Scale bars: 300 μm, 200 μm for insets. (**B**) Venn diagrams summarizing shotgun proteomics data of *S. aureus* (SA) and non–*S. aureus* (NSA) vegetations. (**C**) Venn diagram showing overlap between the total shotgun proteome of SA and NSA vegetations. (**D**) Normalized protein ion abundance plot from the SA and NSA groups showing that the 15 most abundant proteins contributed to over 50% of the total protein abundance (Q1–Q4, proteins in each quartile). (**E**) The rank of the top 15 most abundant proteins in each group (* indicates the same rank in both groups).

and products of inflammatory cells (15, 16), the cumulative infective endocarditis proteome (i.e., combined staphylococcal and non-staphylococcal shotgun data plus additional proteins identified in TAILS, described below) of 1279 gene products was compared with previously published proteomes of human blood (UniProt), blood clots (17), and platelets (18), which showed that 59% (751 products) of the cumulative vegetation proteome overlapped with at least 1 of these proteomes, whereas 41% (528 products) did not (Figure 2B). It should be pointed out that albumin was also detected in vegetations but was filtered out by the applied

mass spectrometry contaminant list (see Supplemental Methods online). The largest overlap was with the platelet proteome, with 241 proteins exclusively shared between platelets and vegetations. Because inflammation was histologically evident in all but 1 of the vegetations (NSA4), we compared the individual vegetation proteomes to the neutrophil proteome (PXD010701) (19) and determined that on average (excepting NSA4), 70% of vegetation proteins were also present in neutrophils (Supplemental Table 3). The proteome of the *Cutibacterium acnes* vegetation (NSA4) conspicuously lacked the abundant acute inflammatory proteins derived from neutrophils and other granulocytes (Supplemental Table 3). The overlap between all the vegetation proteins and the neutrophil proteome accounted for over half of the 528 vegetation proteins that were absent in the other circulation-related proteomes (blood, clot, or platelets) (Figure 2C).

Because most vegetation proteins (82%) are also found in blood, platelets, clots, or neutrophils, we conclude that vegetations are formed de novo from these components. The 230 proteins that were not shared with these proteomes were further characterized by PANTHER annotation to determine their origin (Figure 2D). Most of these were "extracellular region" proteins (Figure 2D), such as immunoglobulins (annotated as extracellular but likely to be circulating), ECM proteins (e.g., collagens, versican, and fibrillin 1), proteases, and their inhibitors (e.g., MMP2, MMP12, Serpin E2), which likely originate in cells invading the vegetations or serum.

To improve proteome depth further following 2-step extraction, 2 vegetations (SA2 and the NSA5 tricuspid valve vegetation, NSA5T) were analyzed by LC-MS/MS following peptide fractionation. The limited amount of vegetation material and the high protein quantity needed for TAILS and fractionation precluded fractionation of all vegetations. Following 2-step protein extraction and trypsin digestion, high-pH reversed-phase chromatography was used to obtain 8 fractions for LC-MS/MS analysis, resulting in 16 LC-MS/MS runs per vegetation (as compared with 2 runs in the nonfractionated analysis). Fractionation resulted in identification of 2139 high-confidence human proteins from these 2 vegetations, with 1060 being found exclusively after fractionation (Figure 3A). Comparison of proteome coverage achieved between these fractionated vegetations, the unfractionated vegetations, and the corresponding vegetation groups is shown in Supplemental Figure 5A. Of the 50% of proteins uniquely identified by fractionation, 36% were found in both fractionated vegetations, and 14% were found in only 1 of the 2 (7% in SA2-F and 6% in NSA5T-F) (Figure 3, A and B). The majority of the proteins uniquely identified after fractionation were of low abundance. Only 7 of the top 150 proteins identified were found exclusively following fractionation, whereas the remaining proteins contributed to the bottom 10th percentile of protein abundance (Figure 3C). This trend was seen also in the proteins exclusively identified in the nonfractionated analysis (Supplemental Figure 5B). Only 2 of the 201 vegetation unique proteins contributed individually to more than 0.1% of the total protein abundance. Both proteins are histone components, histone 2A type 2-C and type 3, which contributed 1.58% and 0.13% of the total protein abundance, respectively. The proteins from fractionated samples were combined with those in the nonfractionated ones, and comparison with the blood, clot, platelet, and neutrophil proteomes was repeated (Figure 3D). This showed a proportionate increase in the vegetation proteins overlapping with the respective proteomes, with 81.4% of the proteins being found in at least 1 of the other proteomes. Secreted proteins accounted for 13 of the 201 additional vegetation-specific proteins that did not match blood, platelet, or neutrophil proteomes. These included the proteoglycan aggrecan, proteases (cathepsin W, trypsin-1, trypsin-3, and F10), protease inhibitors (*TIMP3* and *SERPINA2*), and other matrix-related molecules.

Microbial proteomes identified in vegetations. Mass spectral searches against the proteomes of the respective infectious organisms identified several proteins in the unfractionated vegetations, including some previously unreviewed (Supplemental Data), specifically, 25 *S. aureus* proteins, 2 *S. parasanguinis* proteins, 130 *Candida parapsilosis* proteins, and 714 *C. acnes* proteins (Supplemental Data). An additional 447 *S. aureus* and 5 *S. parasanguinis* proteins were identified following fractionation of the respective vegetations. Several secreted *S. aureus* virulence factors were identified (Supplemental Data), including Panton-Valentine leukocidin and the leukocidins *LukD*, *LukG*, and *LukH*; as well as several proteins involved in iron binding, scavenging, and uptake, including isdB, the clp family of proteases, and siderophore synthesis enzymes in the sbn family. Non-staphylococcal vegetations contained several known virulence factors, including glyceraldehyde-3-phosphate dehydrogenase, endoglycoceramidase (PPA0644 and B1B09_07950), hyaluronate lyase, and others from *C. acnes* (Supplemental Data). The *C. acnes* sample also contained 10 microbe-specific proteins with protease activity, including metalloproteinases and a trypsin-like protease. Many uncharacterized proteins from *C. acnes* and *Candida parapsilosis* were identified, some with as many as 34 unique peptides. No high-confidence peptides from *S. bovis* or *S. dysgalactiae* were identified in their respective vegetations.

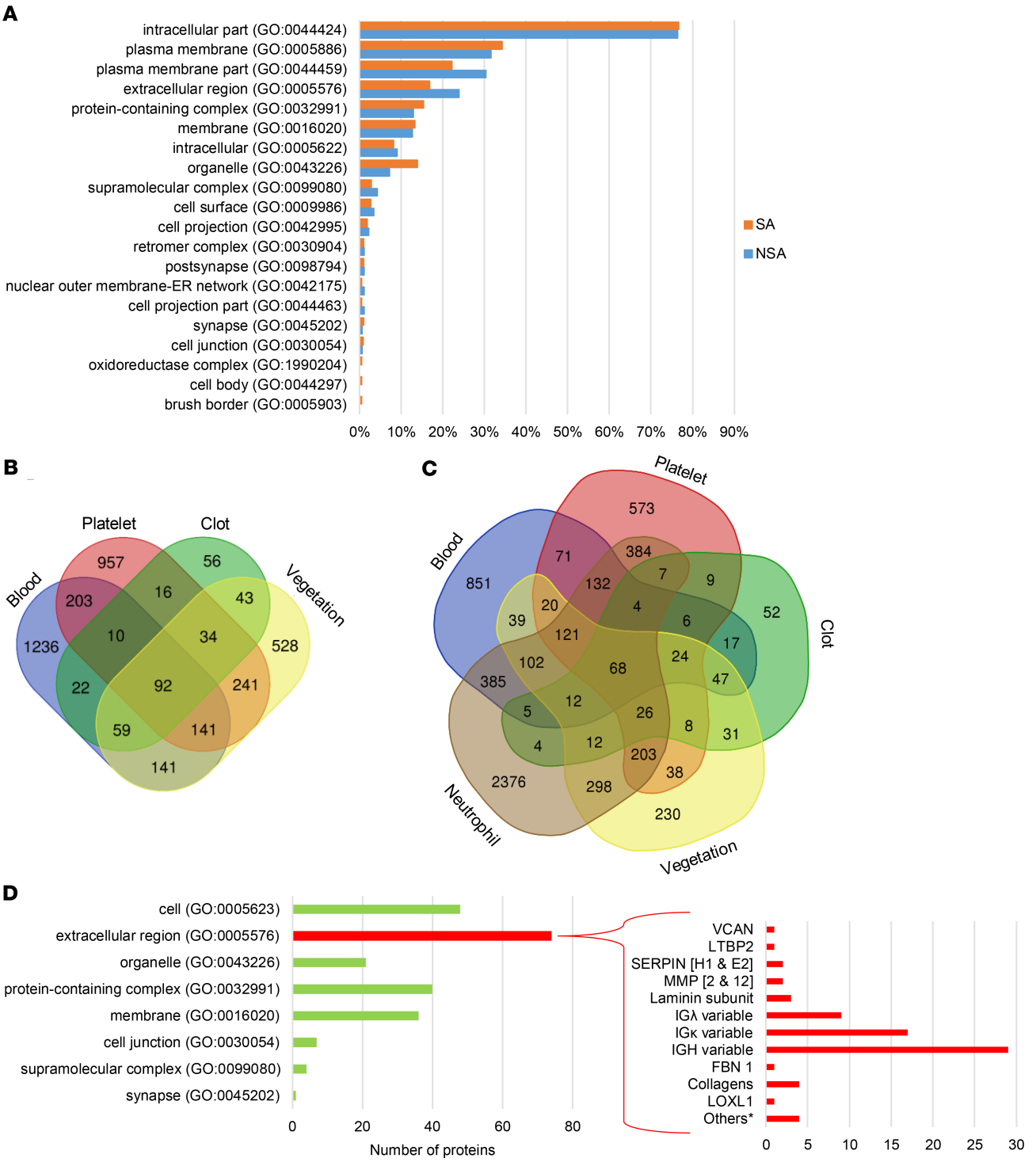


Figure 2. Vegetation proteins predominantly originate in the circulation and inflammatory cells. (**A**) Bioinformatics analysis revealed that human vegetation components of *S. aureus* (SA) and non–*S. aureus* (NSA) groups had largely similar origins. GO, Gene Ontology. (**B**) Venn diagram of the aggregate staphylococcal and non-staphylococcal vegetation proteome overlap with blood, platelet, and clot proteomes. (**C**) Venn diagram of the aggregate staphylococcal and non-staphylococcal vegetation proteome overlap with blood, platelet, clot, and neutrophil proteomes. (**D**) Bioinformatics analysis of vegetation proteins not shared with blood, platelet, clot, or neutrophil proteomes reveals many proteins annotated as extracellular. Some examples are shown at right (* indicates gene products that did not fit into the listed categories, namely *AEBP1*, *PXDN*, *LRRC32*, and *IGFBP2*).

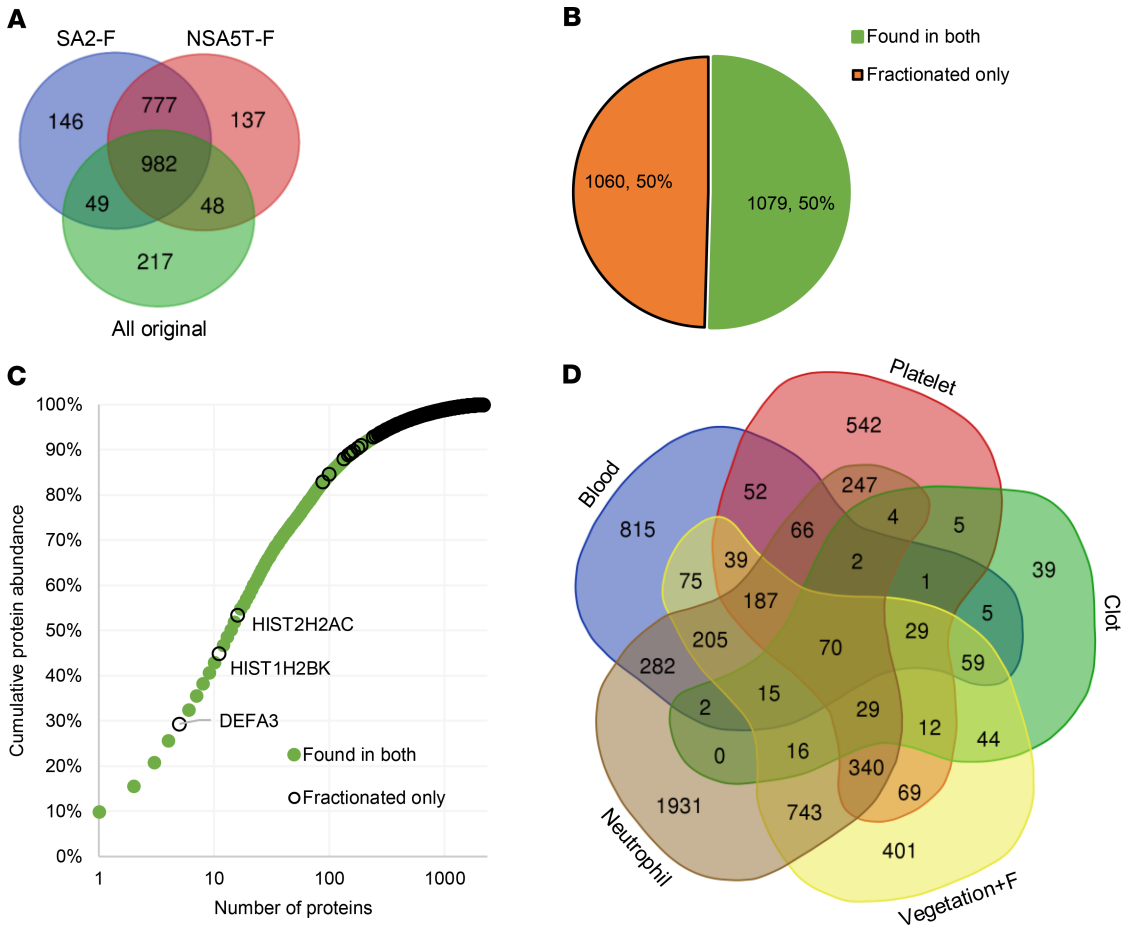


Figure 3. Fractionation increased vegetation proteome coverage by identifying numerous low-abundance proteins. (**A**) Venn diagram showing the overlap between the 2 fractionated vegetations, *S. aureus* 2 fractionated (SA2-F) and non–*S. aureus* 5T fractionated (NSA5T-F), and all the proteins identified in vegetations by single-run analysis. (**B**) Pie chart showing that 50% of the proteins identified in the fractionated vegetations were not previously found in any of the vegetations. (**C**) Normalized protein ion abundance plot from the fractionated vegetations showing the contribution of the fractionation unique proteins (black circles) and the proteins found in the fractionated and nonfractionated vegetations to the overall vegetation proteome. Proteins within the top 80th percentile are labeled by gene symbol. (**D**) Venn diagram of all vegetation proteomes (including after fractionation) with reported blood, platelet, clot, or neutrophil proteomes.

TAILS analysis of vegetations suggests pervasive proteolysis of vegetation components. The TAILS strategy is designed to improve detection of internal protein cleavages by enrichment of protein N-termini. This is achieved by experimental amine labeling at the protein level (i.e., before trypsin digestion before LC-MS/MS) followed by depletion of the tryptic N-termini using a polymer reactive only with unblocked N-termini (20). Because TAILS is effectively also a fractionation method, i.e., it reduces the number of peptides arising from each protein to generate a less complex sample, it can potentially expand proteome coverage (20). Although TAILS searches are innately based on a single peptide per protein that identifies a blocked/labeled N-terminus, only proteins with 2 or more TAILS peptides identified were considered for supplementation of the vegetation proteome. We thus identified an additional 118 human proteins (13 being additional isoforms) that were not detected by shotgun LC-MS/MS using identical search parameters (97 staphylococcal, 68 non-staphylococcal, 42 uniquely present in the staphylococcal group, 21 in the non-staphylococcal group, and 42 found in TAILS analysis of both groups) (Figure 4A).

Dimethyl labeling efficiency was comparable (77% and 79%) for staphylococcal and non-staphylococcal groups (Supplemental Table 2). Compared with shotgun analysis, TAILS provided a 3-fold reduction in the number of unlabeled (trypsin-generated) peptides (6170 in the shotgun and 1808 in the TAILS) and a greater than 2-fold increase in the proportion of N-terminally blocked (dimethylated, acetylated, or cyclized) peptides identified (Figure 4B). The N-terminally blocked internal peptides identified in TAILS and shotgun analysis were combined to derive the total complement of internal cleavages, i.e., the vegetation degradome.

Figure 4. Overview of the vegetation proteome and degradome defined by TAILS. (**A**) Venn diagram comparing proteins identified by at least 2 peptides in staphylococcal (SA) or non-staphylococcal (NSA) TAILS or single-run shotgun (SG) approaches. (**B**) Pie charts of peptides from the shotgun (left) and TAILS (right) with blocked N-termini (dimethyl, acetyl, or pyroglutamate modification as indicated in the key) or tryptic (unlabeled/unblocked) peptides. Blocked N-termini were enriched over 2-fold by TAILS. (**C**) Sunburst plot showing that blocked natural N-termini (blue shades) were preferentially acetylated (yellow, outer ring) or arose from constitutive N-terminal cleavage. These were not included in the degradome. Internal blocked peptides (orange) are typically blocked by dimethyl labeling (gray, outer ring) or cyclization (green, outer ring) but rarely acetylation (yellow, outer ring). (**D**) Venn diagram of TAILS-identified internal N-termini in SA and NSA groups. (**E**) Internal peptides seen in at least 4 of the NSA vegetations and none of the SA vegetations, in at least 2 of the SA vegetations and none of the NSA, or in at least 2 of the SA and 4 of the NSA (Both; 2 and 4), were compared with those seen inconsistently within groups (idiosyncratic).

All N-terminal–blocked/–labeled peptides from the shotgun and TAILS analysis were sorted bioinformatically as either natural or internal. Natural peptides include those blocked at the initiator methionine, those starting 1 amino acid after the initiator methionine, or those immediately following the signal peptide, propeptide, or transit peptide sequences (Figure 4C). Because these peptides result from natural protein processing, they were not included in the vegetation degradome. The majority of natural peptides (57%) were blocked by acetylation, indicating a likely intracellular origin (Figure 4C). The remaining peptides (internal peptides, Figure 4C) likely arise from proteolytic cleavages occurring internally in the protein. Of these, the majority possessed free N-termini blocked experimentally in the TAILS workflow (by reductive dimethylation or by cyclization) (97%) whereas a smaller proportion (3%) were blocked by acetylation (Figure 4C).

Comparing internal peptides identified in staphylococcal and non-staphylococcal vegetations revealed that only 1250 (34%) of the total 3699 N-termini were found in at least 1 sample from both groups (Figure 4D). Instead, it was observed that idiosyncratic internal peptides comprised the majority (Figure 4E). In staphylococcal vegetations, only 503 of the 2099 N-termini/cleavage sites (24%) were present in all 3 vegetations. Non-staphylococcal vegetations shared only 127 of the 2850 N-termini/cleavage sites (4%) between all 5 (with the average overlap in any combination of 3 non-staphylococcal cases being 18%). Additionally, in NSA5, where 2 distinct vegetations from the same patient were analyzed, only 25% (526 of 2120) of N-termini were shared, contrasting with the observed higher compositional similarity of these vegetations. If the cleaved peptides found exclusively in each group were dependent wholly on the microorganism, then one would expect to see consistency between samples of the same group; i.e., the same peptide would be found in at least 2/3 of the vegetations in the staphylococcal group or 4/6 in the non-staphylococcal group (including NSA5 aortic valve and NSA5 tricuspid valve vegetations listed separately). In contrast, visualization of the number of peptides shared between 2 out of 3 vegetations in the staphylococcal group or 4 out of 6 in the non-staphylococcal group revealed greater intergroup similarity than the overall similarities between all samples in each group (Figure 4E). Peptides that did not fall into 1 of these 3 categories (i.e., 2/3 found exclusively in staphylococcal, 4/6 found exclusively in non-staphylococcal, or found in both 2/3 staphylococcal and 4/6 non-staphylococcal) were considered idiosyncratic, meaning they are most likely vegetation-specific or microorganism-specific cleavage sites that may occur infrequently, represent inefficient cleavages, or rely on the abundance of the originating protein, which may vary. Raggedness of N-termini as a consequence of exopeptidase activity could also cause variability between vegetation degradomes. To investigate this, we first sought peptide sequences sharing the first or last 9 amino acids (Supplemental Table 4, e.g., peptide GTGSETESPR shares the same last 9 amino acids as the peptide YGTGSETESPR) from the 3699 internal peptide sequences. Peptides meeting this specification were next filtered to exclude peptides with similar sequences but differing by more than 3 amino acids at each end, which are unlikely to be the result of exopeptidase activity; with missed trypsin cleavages; and with peptide sequence divergence before or after the 9 amino acids (i.e., AEDTAVYYCAR vs. AEDTAVYYCAK). We obtained evidence for exopeptidase activity affecting both the N- and C-termini (Supplemental Table 4). For example, 15 ragged sequences with an average of 7 potential exopeptidase activity events were identified in fibrinogen A (*FGA*) alone (Supplemental Table 4). The number of peptides with N-terminal ragging was 583 (16%) and accounted for 58% of total internal peptide abundance. The number of peptides with C-terminal ragging was 68 (2% of all identified peptides) and accounted for 4% of the total internal peptide abundance. This discrepancy between N- and C-terminal raggedness likely arises from the fact that the TAILS approach we used is designed to detect N-terminal processing. Ragging appears to account for over half (66%, 651 of the 984) of the observed idiosyncratic peptides. NSA4, which showed the least cellular infiltrate, had the lowest number of cleaved peptides.

TAILS identified 1096 proteins in the staphylococcal vegetations with natural or internal N-termini. However, the 1414 internal peptides that were identified originated from only 288 (26.2%) proteins. TAILS identified 1143 proteins in the non-staphylococcal vegetations with natural or internal N-termini. As in the staphylococcal group, the 2145 internal peptides detected in the non-staphylococcal group originated from a fraction of these, i.e., only 374 (32.7%) proteins (Supplemental Table 2). Thus, extensive cleavage of a limited number of proteins was detected in vegetations and possibly reflects the higher abundance of these proteins, their susceptibility to proteases, and/or the presence of proteases that target them. For example, fibrinogen (*FGA*, *FGB*, and *FGG*) and fibronectin had the most numerous internal cleavages, with 249 and 85 unique internal peptides identified, respectively (Figure 5A).

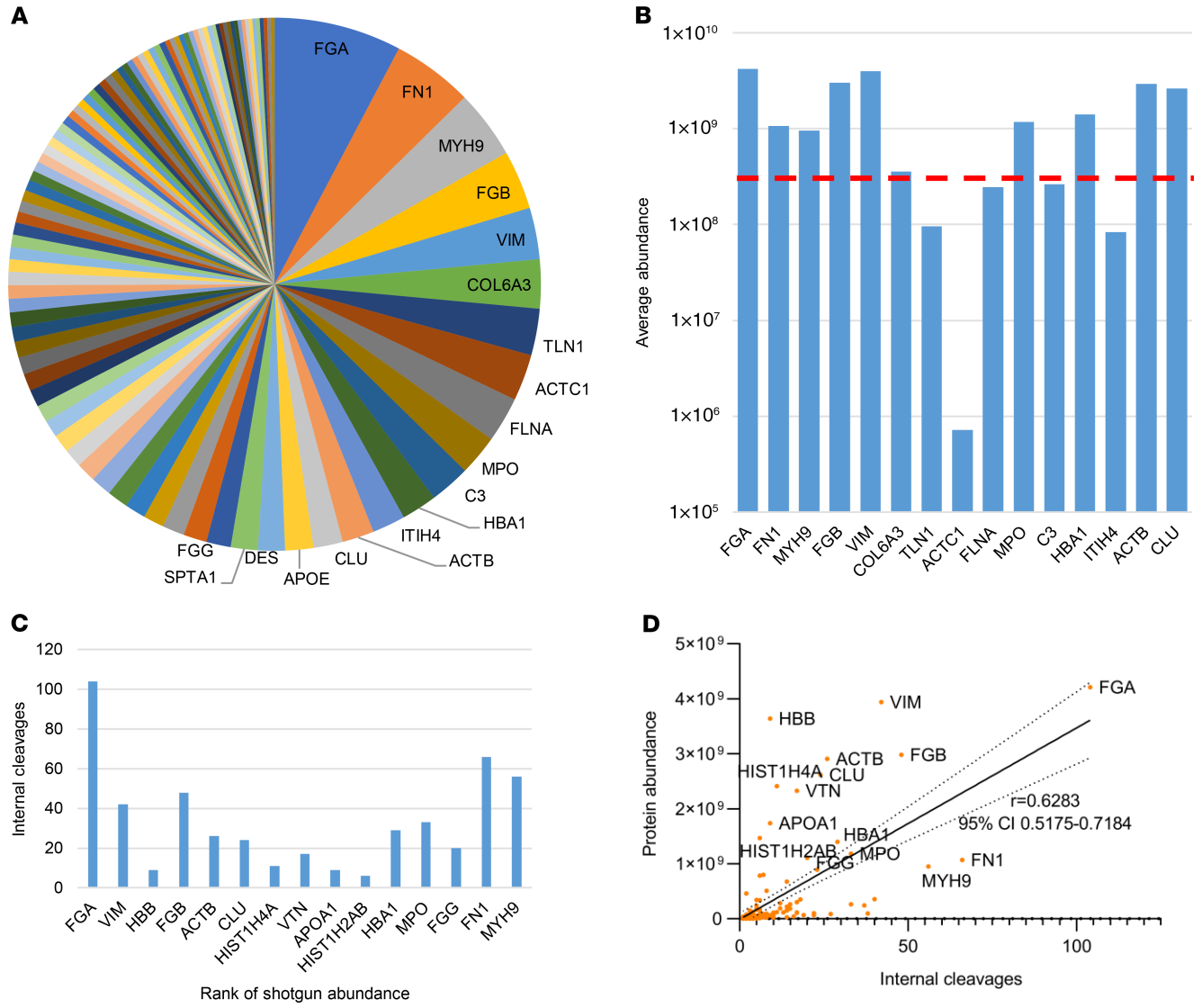


Figure 5. Fibrinogen and fibronectin have the most numerous internal cleavages in vegetations. (**A**) Pie chart of 100 proteins with the most internal N-termini (top 15 are named) shows that fibrinogen (*FGA*, *FGB*, *FGG*) and fibronectin (*FN1*) have the most numerous internal cleavages in vegetations. (**B**) The top 15 most cleaved proteins (from left to right) and their average abundance was compared with overall average protein abundance (dotted red line). (**C**) Number of internal N-termini identified in the top 15 most abundant proteins from the shotgun analysis. (**D**) Pearson correlation plot shows a modest positive correlation (significantly non-0 $P < 0.0001$) between protein abundance and number of internal N-termini. The top 15 most abundant proteins are labeled with their gene symbols.

Although they were among the most abundant proteins, the number of cleavage sites showed only a modest overall correlation with protein abundance (Figure 5, B–D). Comparing internal peptide abundance with relative protein abundance, histone H3.2 had the highest ratio of internal to tryptic peptides identified (for proteins that had both internal and tryptic peptides). The cleavage sites identified in histone H3 matched those known to be produced by granzyme A, which produces a trypsin-like cleavage, and are likely to arise from this serine protease. Notably, neither granzyme A nor histone H3 cleavage peptides were detected in NSA4, where no inflammatory cells were evident histologically.

PICS analysis (21) and WebLogo plots of the internal peptides revealed that the frequency of amino acids in the vicinity of the N-termini was comparable between staphylococcal and non-staphylococcal groups (Supplemental Figure 6, A–C). In addition, there was no discernible difference in the average position of blocked N-termini from each group (Supplemental Figure 6D). The most represented amino acids in internal peptides from both the staphylococcal and non-staphylococcal groups were Arg at P1 and Gln at P1′ and P6′ (Supplemental Figure 6, A and B). In the staphylococcal group, Gln was also

most represented at P3′ to P5′. Overall, the greatest frequency was of Arg at the P1 position, which is explained by the presence of proteases with trypsin-like specificity (Supplemental Table 5) and performance of database searches necessarily with Arg-C specificity because trypsin was used as the working protease for proteomics. Additionally, Gln cyclization can occur, naturally blocking the free N-termini and perhaps enriching cleavages with P1′ Gln more readily in the TAILS workflow.

As illustrative examples of how the TAILS data could be used, identified cleavage sites and previously known sites (annotated by https://www.ebi.ac.uk/merops/search.shtml) in fibrinogen and fibronectin were mapped on their primary structures (Figure 6, A and B). Peptidase cleavages generated by plasmin, thrombin, and neutrophil elastase were included as well as disulfide bonds, alternative sequences, and other known molecular processing events. The TAILS data revealed a plethora of potentially novel cleavage sites; of 152 and 63 cleavages identified in FGA and FGB, respectively, only 24 and 2 were identical to those previously identified in the blood degradome and 5 and 4 with those identified in plasma, respectively (22). We estimate that ragging potentially accounted for up to 124 and 43 peptides from FGA and FGB, respectively, indicating substantial exopeptidase activity.

TAILS plus shotgun data searched against the microbial databases identified 318 unique internal peptides from microbial proteins (Supplemental Data). These included 286 peptides from 152 *C. acnes* proteins, 13 peptides from 11 *Candida parapsilosis* proteins, 8 peptides from 8 *S. parasanguinis* proteins, and 1 peptide from 1 staphylococcal protein. The 2 most highly modified proteins were both from *C. acnes*, i.e., adhesin with 10 unique internal cleavages and uncharacterized protein PPA2210 with 14 internal cleavages. To our knowledge, these microbial degradomes have not been characterized previously in infective endocarditis.

Proteases in vegetations. Although proteases typically have lower abundance than their targets in tissues, we identified 33 human proteases in vegetations without fractionation and an additional 15 following fractionation, including endopeptidases, exopeptidases, and others less well characterized, whose activity could explain the diverse degradomes (Supplemental Table 5). Most proteases were consistently seen in vegetations regardless of the specific pathogen, with 25/33 found with high confidence in at least 5 vegetations and 13/15 found in both fractioned vegetations. The most abundant proteases were plasminogen, prothrombin, and neutrophil elastase. Of additional relevance to prevalence of Arg at the P1 position, cathepsin B, which also has a preference for Arg at P1, was present at relatively high abundance. Complement C1s, complement C1r, and plasma kallikrein were also identified and favor Arg at P1. Proteases produced by neutrophils in addition to neutrophil elastase, such as MMP2, MMP8, and MMP9, were consistently identified.

Potential peptide markers of infective endocarditis. For novel peptides that may reliably reflect the presence of vegetations, outlier internal peptides (ion intensity > [1.5 × IQR] + Q3) that were found in 6 vegetations (at least 2 from the staphylococcal group and 4 from the non-staphylococcal group) were considered. These peptides may be suitable candidates because they were most readily identified and the most consistent between vegetations. Fully tryptic peptides were excluded because they are generated in typical mass spectrometry (MS) workflows and would confound analysis of protein turnover. The remaining peptides were searched individually against the PeptideAtlas database (http://www.peptideatlas.org) (build name: Human Plasma FDR 5% 2010-05) to determine whether they had previously been identified in the circulation (Table 2). This PeptideAtlas build contains a database of peptides identified in human plasma from 91 experiments. Two peptides were selected as preferred candidates on the basis of the search: one from fibronectin and another from complement C3, the parent proteins being abundant components of all vegetations (Table 2 and Supplemental Figure 7). The fibronectin peptide (SVYEQHESTPLR) was present in all the staphylococcal vegetations and in non-staphylococcal vegetations 1, 2, 3, and 5 (found in the tricuspid valve vegetation of NSA5 but not the aortic valve vegetation). The complement C3 peptide (AQMTEDAVDAER) was also found in all the staphylococcal vegetations and in NSA1, -2, -4, and -5 (both vegetations from this case). The fibronectin peptide was reported in only 1 of the 91 serum samples previously analyzed and in low amounts (4 observations/1 million spectra). The complement C3 peptide was not previously identified in human serum. Because the cleaved peptides were found at high abundance in vegetations (ion abundance > [1.5 × IQR] + Q3 of all peptides) and originated in circulating proteins, they and other vegetation-derived peptides may be present in infective endocarditis patient serum, which the present work did not test.

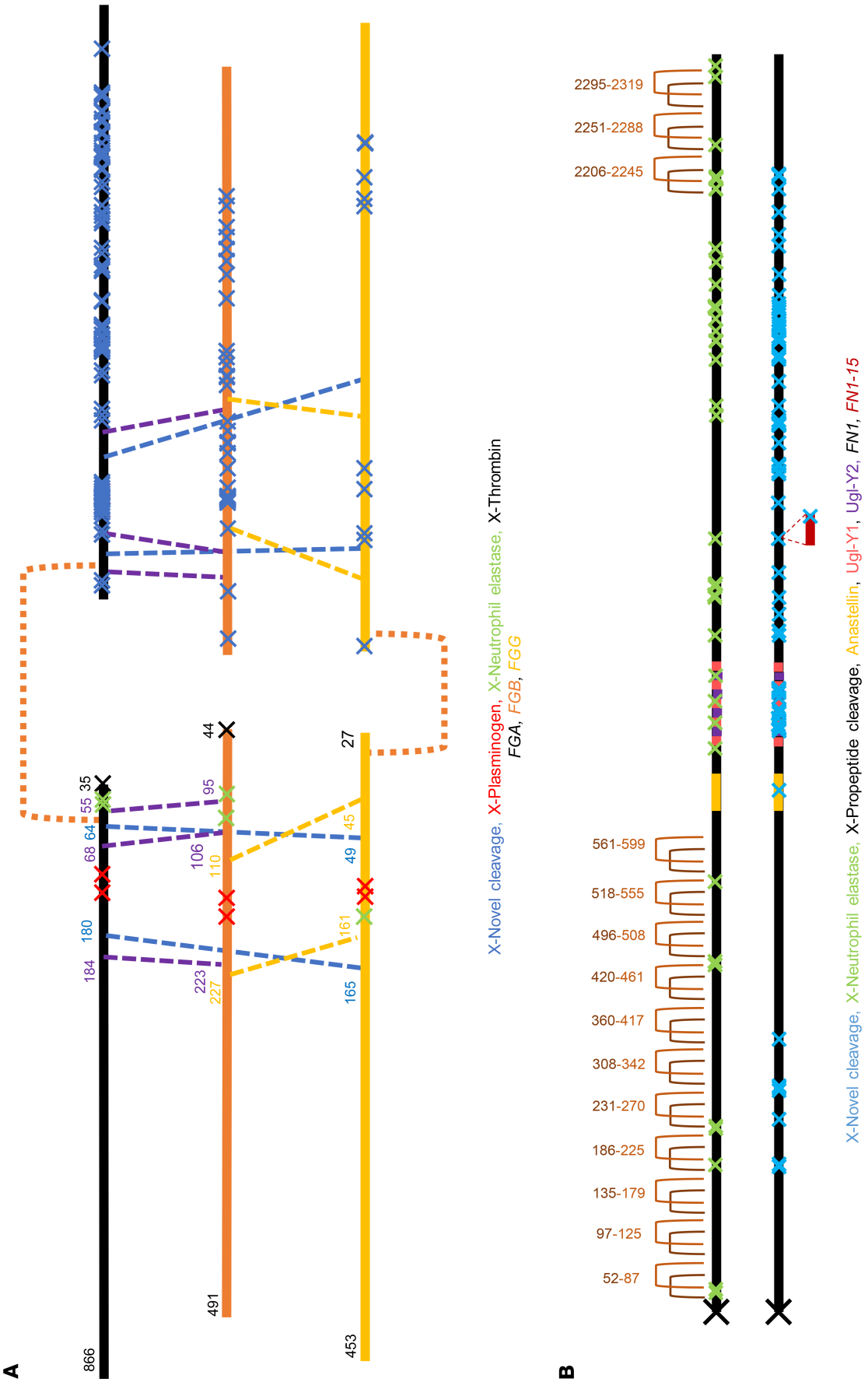

A
866
491
453
35
55
64
68
95
106
110
44
45
49
27
180
184
223
227
161
165
X-Novel cleavage, X-Plasminogen, X-Neutrophil elastase, X-Thrombin
FGA, FGB, FGG
B
2295-2319
2251-2288
2206-2245
561-599
518-555
496-508
420-461
360-417
308-342
231-270
186-225
135-179
97-125
52-87
X-Novel cleavage, X-Neutrophil elastase, X-Propeptide cleavage, Anastellin, Ugl-Y1, Ugl-Y2, FN1, FN1-15

Figure 6. Novel internal N-termini in fibrinogen and fibronectin. (**A**) Fibrinogen complex, showing the 3 component polypeptides *FGA*, *FGB*, and *FGG*. (**B**) Fibronectin dimer. Previously identified N-termini (red, green, or black crosses, annotated at the bottom of each panel by causative protease as appropriate) and novel N-termini (blue crosses) are shown. Disulfide bonds are shown and labeled with their amino acid location, and the isoform names (UniProt) are listed and color-coded.

Discussion

Here, we have defined the proteome and degradome of infective endocarditis vegetations resulting from *S. aureus* or other infections. The infective endocarditis cohort displayed considerable microbial, clinical, and histological diversity and is thus a representative cross section of the infective endocarditis clinical spectrum and of broad relevance to its pathogenesis. A 2-step extraction method maximized the protein yield for subsequent proteomic analysis. The breadth of proteome coverage was expanded by analysis of vegetations from 8 patients and the depth by fractionation of 2 vegetations, and by TAILS, which also reduced sample complexity. Fractionation markedly increased the number of proteins identified, but most of these proteins were of low abundance. The higher abundance proteins (those within the 80th percentile of abundance contribution) that were uniquely identified in the fractionated samples were histone variants and defensin 3. Single peptides from each protein were in fact detected without fractionation, and were abundant, but because the peptides were not unique to that protein, they were not admitted as master proteins in the unfractionated shotgun output. In this case, fractionation provided additional protein unique peptides to fulfill this inclusion criterion of the shotgun analysis. In addition, fractionation of only 2 vegetations nearly doubled the number of proteins identified from single MS runs of 8 vegetations. Efficient extraction and fractionation will thus provide the highest yield of protein identifications in any future vegetation proteome analysis. In this context, the paucity of proteins identified in a previous proteomic analysis of a mycobacterial intracardiac growth may reflect its use of fixed, paraffin-embedded tissue as the protein source, which impairs both protein extraction and MS identification (13).

The vegetation composition elucidated by the present work supports a pathogenic mechanism articulated by Osler, who in his 1885 Gulstonian Lecture on "malignant endocarditis" (10) commented on the histological resemblance of the vegetation to a consolidated thrombus, recognizing the prevalence of fibrin and abundance of platelets ("blood plates of Bizzozero") (Figure 7). Osler mentioned the presence of neutrophils ("white blood corpuscles") and granulation tissue in the deeper regions of vegetations. The 82% overlap between staphylococcal and non-staphylococcal proteomes, similarities in the most abundant proteins, and substantial overlap with blood/serum, platelet, and neutrophil proteomes indicates a dominant host effect in the infective endocarditis pathology. Thus, it might be supposed that similar innate mechanisms and pathways may be activated to challenge an endocardial infection, regardless of the microbial species. Despite the overall similarity between vegetations, however, we observed more consistent proteomes within the staphylococcal group (605/1072 shared, 56%), which presumably elicit a similar immune response, than in the non-staphylococcal group (216/1093 shared, 20%), whose vegetations resulted from diverse microorganisms. Even with any combination of 3 non-staphylococcal samples, the shared proteins were limited to an average of 21%, and 2 vegetations from the same patient (NSA5) demonstrated highly similar proteomes (67%). Thus, while the overall response to an endocardial infection is broadly similar, the infecting microorganism seems to stimulate additional specific responses. Despite the same high-abundance proteins, we observed diverse fibrin deposition patterns, fibronectin staining, and immune cell infiltration histologically. We conclude that vegetation assembly from these components is essentially a random process that could be influenced by many variables, including hemodynamics, clinical history, and treatment.

Surprisingly, vegetation degradomes had greater heterogeneity than the corresponding shotgun proteomes, with an overwhelming prevalence of idiosyncratic cleavages, i.e., cleavage sites that were vegetation specific. It is possible that such cleavages arose from low-abundance proteases that were not detected in the proteomes and their differential access to and interactions with substrates that varied with the diverse spatial arrangement of components such as fibrin, i.e., opportunistic cleavages. Additionally, many of the cleavages could result from ragging, i.e., exopeptidase processing of sites initially arising from endopeptidase activity. Indeed, the estimated ragging events accounted for over half of the idiosyncratic peptides identified. This estimate is necessarily an approximation, assuming that endopeptidases were not responsible for cleavages that generate peptides with 1–3 amino acid differences and that every sequential peptide product of exopeptidase activity was identified. The extent and nature of cellular infiltrate and the different durations of vegetations likely contribute to the number of unshared cleavages between the different cases not explained by ragging.

Table 2. High-abundance proteolytically derived peptides consistently identified in vegetations

Peptide sequence	SA	NSA	Previously identified (based on PeptideAtlas)	Gene name	Accession number [position in master protein]
EEVSGNVSPGTR[A]	3	5	Chymotrypsin-digested liver, ovary, testis, spleen	*FGA*	P02671 [414-425]
VLSGGTTMYPGIADR	3	6	In HeLa cell proteome (HCD fragmentation)	*ACTC1*[B]	P68032 [300-314]
GEFVSETESR[A]	3	5	Serum, human plasma	*FGA*	P02671 [538-547]
LGEFVSETESR[A]	3	4	Serum, human plasma	*FGA*	P02671 [537-547]
SYGTGSETESPR[A]	3	6	Serum, human plasma	*FGA*	P02671 [276-287]
TENGGWTVIQNR[A]	2	4	Human plasma	*FGB*	P02675 [274-285]
STGTWNPGSSER[A]	3	4	Human plasma	*FGA*	P02671 [356-367]
EASILTHDSSIR[A]	2	6	Human plasma	*FGG*	P02679-2 [123-134]
RDNIQGITKPAIR	2	4	Liver	*Hist1H4*	P62805 [24-36]
DEVGGEALGR	3	4	Liver, HCC, dental pulp, salivary gland, and kidney	*HBB*	P68871 [22-31]
SGGTTMYPGIADR	3	5	Similar N-terminal sequences in chymotrypsin-digested liver, ovary, testis, and spleen	*ACTC1*[B]	P68032 [302-314]
GEYGAEALER	3	5	Previously unidentified	*HBA1*	P69905 [23-32]
GAHAGEYGAEALER	3	5	Previously unidentified	*HBA1*	P69905 [19-32]
GQVITIGNER	3	5	Previously unidentified	*ACTC1*[B]	P68032 [247-256]
GKVGAHAGEYGAEALER	2	6	Previously unidentified	*HBA1*	P69905 [16-32]
AHAGEYGAEALER	3	5	Previously unidentified	*HBA1*	P69905 [20-32]
SVYEQHESTPLR[A]	3	4	Low levels identified in human plasma and microparticles	*FN1*	P02751 [1340-1351]
AQMTEDAVDAER	3	5	Previously unidentified	*C3*	P01024 [988-999]
TVLSGGTTMYPGIADR	3	6	Previously unidentified	*ACTB*[B]	P60709 [297-312]
STSVTSDNLSR	3	4	Serum, platelets, many cell lines and tissues	*MAPRE1*	Q15691 [7-17]
TVGSLAGQPLQER	3	4	Serum, HCC cells, cerebrospinal fluid	*APOE*	P02649 [212-224]

Peptide sequences were individually searched against the PeptideAtlas database to determine if they were previously identified and which specific tissue/cell type they were identified in. Underlined peptide sequences were chosen as the preferred candidate peptides. [A]Protein also associated with ECM. [B]More than one gene name corresponded to this peptide. SA, the number of *S. aureus* vegetations containing that peptide ($N = 3$); NSA, the number of non–*S. aureus* vegetations containing that peptide ($N = 6$); HCD, higher energy collisional dissociation; HCC, hepatocellular carcinoma.

Virulence factors mediate tissue adherence, host infiltration, immune resistance/evasion, and dynamic stress responses and confer enhanced pathogen survival, proliferation, and host invasion in animal models of infective endocarditis (12, 23). *S. aureus* has the best characterized proteome of the pathogens in our study, and although we identified only a fraction of its 2000-plus proteins, several virulence factors were detected, including bicomponent pore-forming toxins (24, 25). These included the Panton-Valentine leucocidin (PVL) (26), a pore-forming toxin that induces neutrophil death (26, 27) and is present in 36% of all *S. aureus* isolates and 48% of methicillin-resistant strains in the United States (28). We also identified the 2-component leukotoxin LukED, comprising LukS-H and LukF-G, which synergizes with PVL in neutrophil lysis (29) (the LukS-H component also has specificity toward innate immune cells, binding to T lymphocyte chemokine receptor CCR5; refs. 30, 31), and a γ-toxin expressed by most *S. aureus* strains that lyses erythrocytes and leukocytes (32). Experimental endocarditis models demonstrate the dynamic relationship between *S. aureus* toxin production and endocarditis disease burden. Toxin expression is vital to endocarditis development (33), and increasing toxin production increases virulence (34), although supraphysiological levels of secretion may reduce endocarditis burden and colony size (35). Also identified were several proteins essential to iron scavenging from hemoglobin, including isdB and clpP (36, 37). Although the bloodstream represents a high-iron environment, *S. aureus* colonies within vegetations likely require complex iron-scavenging mechanisms — specifically the production of siderophores, via the siderophore synthesis enzymes (sbnB, sbnE, sbnF, sbnG), which were also identified in our samples (38). Last, the chaperone DnaK and elongation factor Tu, which are associated with bacteremia-inducing *S. aureus* strains, but not noninvasive *S. aureus* strains, were also identified (39). DnaK plays a vital role in the *S. aureus* stress response, and its loss reduces *S. aureus* adhesion to endothelium and biofilm formation (40).

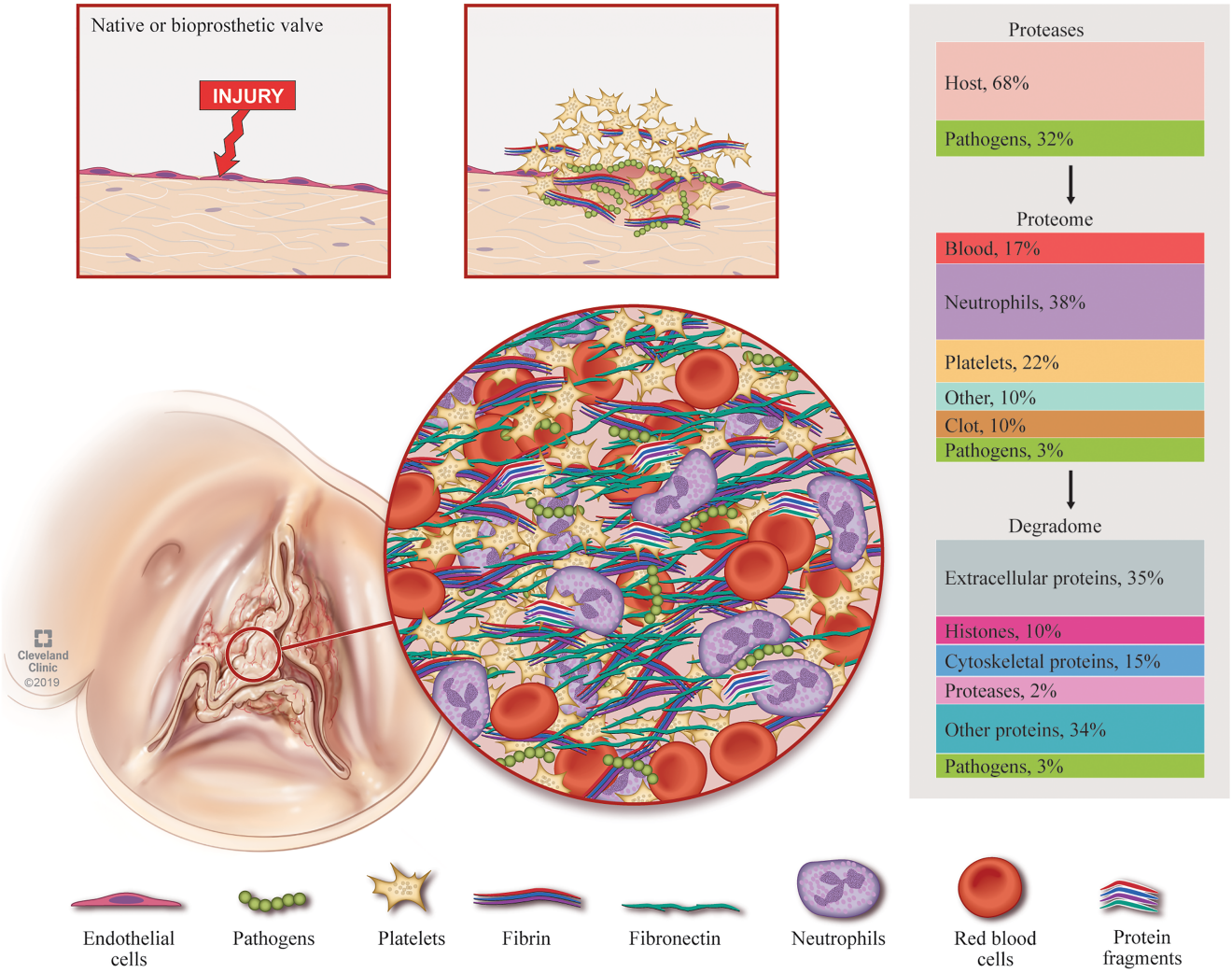


Figure 7. The vegetation proteome and its origins. Native or bioprosthetic valves are infiltrated directly following bacteremia, or valvular damage, exposing prothrombogenic components to create a platelet-rich nidus. The vegetation grows via layering of components starting with opsonization of pathogens by circulating immunoglobins, complement, and other factors of the innate immune system in response to pathogen virulence factors. Circulating fibrin and fibronectin trap red blood cells and recruit additional platelets as they bind to the nidus. The accumulated platelets provoke additional fibrin deposition in the region. Neutrophils invade the site of infection. Inaccessible pathogens within the valve mediate persistent neutrophil chemotaxis via degranulation and NETosis, releasing citrullinated DNA and proteases. NETosis and platelet degranulation further stimulates cellular and protein deposition. Vegetations are thus a conglomerate of multiple proteomes whose respective percentage contributions can vary widely between vegetations. Pathogen and host protease-mediated vegetation turnover may destabilize the vegetation, which, when combined with turbulent valve function, could promote embolization, leading to cerebral or peripheral abscesses, infarcts, and mycotic aneurysms. Human as well as microbial proteases can contribute to protein turnover in vegetations. TAILS analysis suggested extensive turnover of extracellular proteins, with fibrin and fibronectin among the top proteins modified by proteolysis.

Together, these *S. aureus* proteins represent danger markers with strong innate host tropism; i.e., after the elimination of living bacteria, they continue to provoke immune cell influx and thus growth and remodeling of vegetations. Their presence may represent increased production by *S. aureus* in established infective endocarditis vegetations, demonstrating their impact on survival. Despite evidence showing a minimal burden of *S. aureus* on histology, and low abundance of *S. aureus* peptides in our samples, the identification of virulence factors reveals their important role for *S. aureus* pathogenesis in human infective endocarditis.

Although many *C. acnes* and *Candida parapsilosis* proteins were identified in the respective vegetations, the proteomes of both pathogens are largely unreviewed. *C.* (formerly *Propionibacterium*) *acnes* is a ubiquitous skin commensal that is an opportunistic pathogen and a rare cause of prosthetic valve endocarditis (41, 42). *C. acnes* rapidly forms biofilms, which enable adherence to prosthetic material (43). The *C. acnes* vegetation (NSA4) was composed nearly entirely of bacteria and serum components, with few to no immune cells.

The clinical course of *C. acnes* endocarditis is often long and indolent and may lack symptoms of systemic infection — likely due to the slow growth of *C. acnes* and predominant biofilm formation (44). Whole-genome sequencing and in vitro proteomics determined that pathogenic *C. acnes* strains differ from benign ones in virulence factor expression (43, 45–47). We identified some of these putative virulence factors, including the cohemolytic Christie-Atkins-Munch-Petersen factors, cutinase, triacylglycerol lipase, glyceraldehyde-3-phosphate dehydrogenase, and adhesion molecules (43, 46, 47). One of the most abundant *C. acnes* proteins in our analysis, dermatan-binding protein PA-5541, is a potential immunoreactive surface protein and is likely a microbial surface component recognizing adhesive matrix molecule that is highly expressed in pathogenic *C. acnes* strains (47, 48). The abundant *C. acnes* proteome identified in this infective endocarditis vegetation implies robust colonization, making a strong case for primary pathogenicity of this commensal.

TAILS provided a cumulative history of proteolysis and proteome turnover in vegetations (Figure 7). Previously, select host degradative enzymes (MMPs and neutrophil elastase) were investigated in infective endocarditis vegetations (49), but a proteome-level analysis of proteases in vegetations and vegetation component proteolysis was unavailable. The TAILS data suggest that although vegetation proteins derive mostly from deposited circulatory cells and proteins, proteolytic products may be formed in situ. For example, comparison of the fibrin cleavage products identified here with published N-terminomics of whole blood and plasma showed only a 14% overlap between the 2 data sets (22, 50). Along with consistent detection of proteases of broad specificity, such as neutrophil elastase and MMPs, in the vegetations, the numerous proteolytic sites distinct from those attributed to coagulation suggests ongoing degradation in infective endocarditis vegetations by vegetation-intrinsic proteases of host/microbial origin (Figure 7), which has several implications for infective endocarditis. Proteolysis may generate new proteoforms that favor or suppress the infection, present danger signals to immune cells, or modify the immune response, contributing to vegetation growth. Proteolytically truncated proteoforms may aggregate and enable consolidation of the vegetation, contributing to its longevity and persistence. Alternatively, proteolysis of abundant components could destabilize the vegetation, facilitate vegetation fragmentation by hemodynamic forces, and lead to embolism. These effects likely occur concurrently in vegetations, contributing to uncertain clinical behavior, but may ultimately be deleterious to patients.

The proteomics data aligned with the clinical and histological pictures. For example, patient NSA4 had a prolonged course, with subclinical symptoms for approximately 8 months until echocardiography disclosed a dysfunctional prosthetic aortic valve and a large vegetation. Given concern for an atypical organism, blood cultures were held for 10 days and eventually grew *C. acnes*. The patient was placed on appropriate antibiotics and underwent surgery 5 days later. This short period between antibiotic initiation and surgery likely accounts for the large bacterial burden in the vegetation but does not explain the lack of immune cell infiltration. A contrasting case was SA1, who developed severe chest pain, fever to 39.4°C, and tachycardia a day after cocaine injection. After 3 days of symptoms, the patient was found to have methicillin-resistant *S. aureus* bacteremia. Transesophageal echocardiography showed severe aortic insufficiency, vegetations on aortic and mitral valves, an aortic anular abscess, and a fistula tract connecting the aortic sinus and left atrium. The patient was given antibiotics and underwent surgery 12 days following antibiotic administration. Relatively few *S. aureus* proteins were found in the tissue. In contrast, abundant inflammatory cell proteins, including histones, proteases, and neutrophil defense proteins, were identified, together suggesting a robust host response and a microbicidal effect of early antibiotics.

Diagnosis of infective endocarditis relies on clinical suspicion, based on the patient's history and risk factors (intravenous drug use, prosthetic valve implantation, or previous diagnosis of infective endocarditis). The most frequent symptom is fever, which lacks specificity, and the only laboratory diagnostic test in the modified Duke criteria is blood culture. Hence, delayed diagnosis is common and diagnostic biomarkers would be helpful. Peptides identified here could provide multiplex/multiple reaction monitoring assays as markers for vegetation formation, growth, or embolism because contemporary MS techniques can yield results quickly following sample collection. Two potential peptide biomarkers we identified originated from serum proteins, which may increase their accessibility and diagnostic value. The fibronectin peptide was previously associated with microparticles in circulation, which are induced during the vascular stress response (51). The complement C3 peptide was previously unidentified in human serum, even in proteomics performed following depletion of high-abundance proteins (52). Additional studies using blood from patients known to have a vegetation and those without (including patients with only bacteremia or deep vein thrombosis) are needed to validate these and other peptides as potential diagnostic infective endocarditis biomarkers. The 2 peptides selected here were

consistently abundant in vegetations, originated in serum proteins, and had not been reported in analysis of noninfective endocarditis serum.

Infective endocarditis necessitates prolonged antibiotic therapy and requires surgery in up to 50% of cases (4, 7). Surgery is undertaken to prevent potential embolization, to remedy valvular dysfunction as a result of vegetations or leaflet destruction, or for source control in the setting of recalcitrant organisms or prolonged bacteremia (4). Medical adjuvants to antibiotics that prevent vegetation growth or reduce the risk of embolization could reduce morbidity and possibly the need for surgery. Given the important role for platelets in vegetation formation, prior studies have investigated the use of aspirin for embolism prevention. While retrospective analysis demonstrated promising results (53), a randomized trial trended toward increased bleeding risk without reduction in embolization (54). Other preclinical studies have investigated anticoagulants (55), DNase I (56), and specific virulence factor–directed therapies (25). The current study offers support for additional therapeutic possibilities. In an established vegetation, abundant proteolysis appears to be balanced by ongoing deposition and thrombosis. If the balance tilts toward increased proteolysis or reduced deposition, the vegetation burden may be reduced, and bacterial colonies may become better exposed to antibiotic killing and immune clearance. Additionally, targeting the identified pathogen virulence factors may disrupt the immune recruitment and intense response, also likely moderating vegetation growth.

Significant sample and clinical variability, including variable preoperative antibiotic treatment times, diverse organisms, and heterogeneous vegetation locations (including left- and right-heart vegetations), is inevitable in an analysis such as this and could be considered a limitation. In addition, the inherent heterogeneity of the non-staphylococcal group, in addition to clinical variability between all samples, dictated the inevitability that a quantitative comparison between the non-staphylococcal and staphylococcal groups would be statistically unreliable. However, the diversity is clinically representative and provided a comprehensive systems-level view and a detailed molecular repertoire of infective endocarditis vegetations for the first time to our knowledge. This study has also identified a plethora of unreviewed microbial proteins and novel N-terminal microbial peptides. Additionally, the staphylococcal group proteomics data can be reliably compared to a cohort of vegetations resulting from another infectious organism in future studies and each of our non-staphylococcal vegetations data sets to other studies with the represented microorganisms. Despite the variability between samples, we have shown a major contribution of proteolysis within all vegetations analyzed. Furthermore, proteolytic remodeling of the components is potentially informative about mechanisms of embolization. These vegetation proteomes and degradomes add modern molecular biology detail to Osler's scholarly understanding from an earlier era and provide insights on innate mechanisms that attempt to contain an intravascular microbial infection.

Methods

Patients and vegetations. The patient cohort ranged in age from 32 to 78 years (average age 57 ± 15 years) and included 4 females and 4 males (Table 1). Two vegetations occurred on bioprosthetic valves, 5 on native valves, and 1 at the edge of a right atrial fistula. Three patients with methicillin-resistant *S. aureus* infections formed the *S. aureus* (staphylococcal) cohort. Five patients infected by *Streptococcus* strains (*S. bovis*, *S. parasanguinis*, and *S. dysgalactiae*), *C. acnes*, and *Candida parapsilosis* (Table 1, Supplemental Table 1) were grouped as the non–*S. aureus* (non-staphylococcal) cohort.

Protein analysis by MS. Proteins were extracted from staphylococcal and non-staphylococcal vegetations using a 2-step extraction method and prepared for shotgun proteomics and TAILS with or without further fractionation as described in the Supplemental Methods. Individual LC-MS/MS raw files were searched against human and pathogen databases (UniProt) using Proteome Discoverer 2.2 (Thermo Fisher Scientific). Search parameters and additional details are specified in the Supplemental Methods. A 1% FDR was applied to identify high-confidence proteins.

Statistics. Peptides were identified using a precursor mass tolerance of 10 ppm and fragment mass tolerance of 0.6 Da. Peptides were validated using an FDR of 1% against a decoy database. Chromatographic retention time alignment was used across like (e.g., all TAILS) samples for accurate label-free quantitation comparison and improved peptide identifications. Only high-confidence proteins (containing peptides at a 99% confidence level or higher) were recorded from each sample for data analysis. Shotgun data required a minimum of 2 high-confidence peptides for protein identification, and TAILS (by definition) required a single peptide. Statistical analyses were performed in either GraphPad Prism 8 or Microsoft Excel 2013. Pearson correlations tests were performed in GraphPad Prism to determine if the slope

was significantly non-0, P value < 0.01, and the r value with a 95% confidence interval. Excel was used to perform outlier tests for normally distributed, a 2-sided Grubbs' outlier test, and for skewed data, high outlier > Q3 + (1.5 × IQR) or low outlier < Q1 – (1.5 × IQR). Data visualizations were created using CLIPPICS (http://clipserve.clip.ubc.ca/pics/), WebLogo (https://weblogo.berkeley.edu/logo.cgi), GraphPad Prism 8, and Microsoft Excel. Venn diagrams were created using http://bioinformatics.psb.ugent.be/webtools/Venn/. The UpSet plot was generated in R 4.0.0 using the UpSetR package, and the dumbbell plot was generated in R 4.0.0 using the Plotly package (57, 58). A P value less than 0.05 was considered significant in all statistical tests.

Study approval. Vegetations were collected prospectively during open-heart surgery for infective endocarditis under an approved Cleveland Clinic Institutional Review Board protocol (protocol 16-1521), and patient samples were obtained with verbal patient consent and a printed research information sheet. Neither genetic testing nor patient-identifiable research was performed, and tissue samples collected were those otherwise discarded during surgery. The Cleveland Clinic Institutional Review Board reviewed the study goals and its methods and waived written patient consent.

Author contributions

DRM, JCW, BBW, and SSA designed experiments. DRM and JCW performed experiments. DRM, JCW, CDT, ERR, and DES analyzed the data. DRM, JCW, EHB, GBP, and SSA wrote the paper.

Acknowledgments

This work was supported by the Allen Distinguished Investigator Program, through support made by The Paul G. Allen Frontiers Group and the American Heart Association (to SSA). The Fusion Lumos instrument was purchased via an NIH shared instrument grant, 1S10OD023436-0.

Address correspondence to: Suneel S. Apte, Biomedical Engineering-ND20, Cleveland Clinic Lerner Research Institute, 9500 Euclid Avenue, Cleveland, Ohio 44195, USA. Phone: 216.445.3278; Email: aptes@ccf.org.

1. Pant S, et al. Trends in infective endocarditis incidence, microbiology, and valve replacement in the United States from 2000 to 2011. *J Am Coll Cardiol*. 2015;65(19):2070–2076.
2. Toyoda N, Chikwe J, Itagaki S, Gelijns AC, Adams DH, Egorova NN. Trends in infective endocarditis in California and New York State, 1998-2013. *JAMA*. 2017;317(16):1652–1660.
3. Slipczuk L, et al. Infective endocarditis epidemiology over five decades: a systematic review. *PLoS One*. 2013;8(12):e82665.
4. AATS Surgical Treatment of Infective Endocarditis Consensus Guidelines Writing Committee Chairs, et al. 2016 The American Association for Thoracic Surgery (AATS) consensus guidelines: Surgical treatment of infective endocarditis: executive summary. *J Thorac Cardiovasc Surg*. 2017;153(6):1241–1258.e29.
5. Hussain ST, Shrestha NK, Gordon SM, Houghtaling PL, Blackstone EH, Pettersson GB. Residual patient, anatomic, and surgical obstacles in treating active left-sided infective endocarditis. *J Thorac Cardiovasc Surg*. 2014;148(3):981–8.e4.
6. Lalani T, et al. In-hospital and 1-year mortality in patients undergoing early surgery for prosthetic valve endocarditis. *JAMA Intern Med*. 2013;173(16):1495–1504.
7. Tleyjeh IM, et al. A systematic review of population-based studies of infective endocarditis. *Chest*. 2007;132(3):1025–1035.
8. Tleyjeh IM, Kashour T, Zimmerman V, Steckelberg JM, Wilson WR, Baddour LM. The role of valve surgery in infective endocarditis management: a systematic review of observational studies that included propensity score analysis. *Am Heart J*. 2008;156(5):901–909.
9. Ely D, et al. Histological findings in infective endocarditis. *Open Forum Infect Dis*. 2016;3(suppl 1):1111.
10. Osler W. The Gulstonian Lectures, on malignant endocarditis. *Br Med J*. 1885;1(1262):467–470.
11. Thomas S, Liu W, Arora S, Ganesh V, Ko YP, Höök M. The complex fibrinogen interactions of the *Staphylococcus aureus* coagulases. *Front Cell Infect Microbiol*. 2019;9:106.
12. Foster TJ, Geoghegan JA, Ganesh VK, Höök M. Adhesion, invasion and evasion: the many functions of the surface proteins of *Staphylococcus aureus*. *Nat Rev Microbiol*. 2014;12(1):49–62.
13. Sass LA, et al. A 1-year-old with *Mycobacterium tuberculosis* endocarditis with mass spectrometry analysis of cardiac vegetation composition. *J Pediatric Infect Dis Soc*. 2016;5(1):85–88.
14. Kockmann T, Carte N, Melkko S, Keller UAD. Identification of protease substrates in complex proteomes by iTRAQ-TAILS on a Thermo Q Exactive instrument. In: Grant J, Li H, eds. *Analysis of Post-Translational Modifications and Proteolysis in Neuroscience*. Humana Press; 2015:187–207.
15. Holland TL, et al. Infective endocarditis. *Nat Rev Dis Primers*. 2016;2:16059.
16. Hollanders G, De Scheerder I, De Buyzere M, Ingels G, Bogaert S, Clement DL. A six years review on 53 cases of infective endocarditis: clinical, microbiological and therapeutical features. *Acta Cardiol*. 1988;43(2):121–132.
17. Stachowicz A, et al. Optimization of quantitative proteomic analysis of clots generated from plasma of patients with venous

thromboembolism. *Clin Proteomics*. 2017;14:38.
18. Prudova A, Serrano K, Eckhard U, Fortelny N, Devine DV, Overall CM. TAILS N-terminomics of human platelets reveals pervasive metalloproteinase-dependent proteolytic processing in storage. *Blood*. 2014;124(26):e49–e60.
19. Grabowski P, et al. Proteome analysis of human neutrophil granulocytes from patients with monogenic disease using data-independent acquisition. *Mol Cell Proteomics*. 2019;18(4):760–772.
20. Kleifeld O, et al. Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. *Nat Protoc*. 2011;6(10):1578–1611.
21. Schilling O, auf dem Keller U, Overall CM. Factor Xa subsite mapping by proteome-derived peptide libraries improved using WebPICS, a resource for proteomic identification of cleavage sites. *Biol Chem*. 2011;392(11):1031–1037.
22. Wildes D, Wells JA. Sampling the N-terminal proteome of human blood. *Proc Natl Acad Sci U S A*. 2010;107(10):4561–4566.
23. Brinkman CL, Patel R. Molecular pathogenesis of infective endocarditis. In: Sails A, Tang YW, eds. *Molecular Medical Microbiology*. Academic Press; 2015:811–822.
24. Becher D, et al. A proteomic view of an important human pathogen--towards the quantification of the entire *Staphylococcus aureus* proteome. *PLoS One*. 2009;4(12):e8176.
25. Alonzo F, Torres VJ. The bicomponent pore-forming leucocidins of *Staphylococcus aureus*. *Microbiol Mol Biol Rev*. 2014;78(2):199–230.
26. Genestier AL, et al. *Staphylococcus aureus* Panton-Valentine leukocidin directly targets mitochondria and induces Bax-independent apoptosis of human neutrophils. *J Clin Invest*. 2005;115(11):3117–3127.
27. Löffler B, et al. *Staphylococcus aureus* Panton-Valentine leukocidin is a very potent cytotoxic factor for human neutrophils. *PLoS Pathog*. 2010;6(1):e1000715.
28. Brown ML, et al. Prevalence and sequence variation of Panton-Valentine leukocidin in methicillin-resistant and methicillin-susceptible *Staphylococcus aureus* strains in the United States. *J Clin Microbiol*. 2012;50(1):86–90.
29. Ventura CL, et al. Identification of a novel *Staphylococcus aureus* two-component leukotoxin using cell surface proteomics. *PLoS One*. 2010;5(7):e11634.
30. Alonzo F, et al. CCR5 is a receptor for *Staphylococcus aureus* leukotoxin ED. *Nature*. 2013;493(7430):51–55.
31. Yoong P, Torres VJ. The effects of Staphylococcus aureus leukotoxins on the host: cell lysis and beyond. *Curr Opin Microbiol*. 2013;16(1):63–69.
32. Foster TJ, Geoghegan JA. Chapter 37 - Staphylococcus aureus. In: Tang YW, Sussman M, Liu D, Poxton I, Schwartzman J, eds. *Molecular Medical Microbiology*. Academic Press; 2015:655–674.
33. Cheung AL, et al. Diminished virulence of a sar-/agr- mutant of Staphylococcus aureus in the rabbit model of endocarditis. *J Clin Invest*. 1994;94(5):1815–1822.
34. Pérez-Montarelo D, et al. Pathogenic characteristics of *Staphylococcus aureus* endovascular infection isolates from different clonal complexes. *Front Microbiol*. 2017;8:917.
35. Bayer AS, Ramos MD, Menzies BE, Yeaman MR, Shen AJ, Cheung AL. Hyperproduction of alpha-toxin by Staphylococcus aureus results in paradoxically reduced virulence in experimental endocarditis: a host defense role for platelet microbicidal proteins. *Infect Immun*. 1997;65(11):4652–4660.
36. Farrand AJ, Reniere ML, Ingmer H, Frees D, Skaar EP. Regulation of host hemoglobin binding by the *Staphylococcus aureus* Clp proteolytic system. *J Bacteriol*. 2013;195(22):5041–5050.
37. Mazmanian SK, et al. Passage of heme-iron across the envelope of *Staphylococcus aureus*. *Science*. 2003;299(5608):906–909.
38. Perry WJ, et al. *Staphylococcus aureus* exhibits heterogeneous siderophore production within the vertebrate host. *Proc Natl Acad Sci U S A*. 2019;116(44):21980–21982.
39. Liew YK, Awang Hamat R, van Belkum A, Chong PP, Neela V. Comparative exoproteomics and host inflammatory response in Staphylococcus aureus skin and soft tissue infections, bacteremia, and subclinical colonization. *Clin Vaccine Immunol*. 2015;22(5):593–603.
40. Singh VK, Syring M, Singh A, Singhal K, Dalecki A, Johansson T. An insight into the significance of the DnaK heat shock system in Staphylococcus aureus. *Int J Med Microbiol*. 2012;302(6):242–252.
41. Banzon JM, Rehm SJ, Gordon SM, Hussain ST, Pettersson GB, Shrestha NK. Propionibacterium acnes endocarditis: a case series. *Clin Microbiol Infect*. 2017;23(6):396–399.
42. van Valen R, de Lind van Wijngaarden RA, Verkaik NJ, Mokhles MM, Bogers AJ. Prosthetic valve endocarditis due to Propionibacterium acnes. *Interact Cardiovasc Thorac Surg*. 2016;23(1):150–155.
43. Achermann Y, Goldstein EJ, Coenye T, Shirtliff ME. Propionibacterium acnes: from commensal to opportunistic biofilm-associated implant pathogen. *Clin Microbiol Rev*. 2014;27(3):419–440.
44. Clayton JJ, Baig W, Reynolds GW, Sandoe JAT. Endocarditis caused by Propionibacterium species: a report of three cases and a review of clinical features and diagnostic difficulties. *J Med Microbiol*. 2006;55(pt 8):981–987.
45. McDowell A, et al. An expanded multilocus sequence typing scheme for propionibacterium acnes: investigation of 'pathogenic', 'commensal' and antibiotic resistant strains. *PLoS One*. 2012;7(7):e41480.
46. Brüggemann H, et al. The complete genome sequence of Propionibacterium acnes, a commensal of human skin. *Science*. 2004;305(5684):671–673.
47. Holland C, et al. Proteomic identification of secreted proteins of Propionibacterium acnes. *BMC Microbiol*. 2010;10:230.
48. Lodes MJ, et al. Variable expression of immunoreactive surface proteins of Propionibacterium acnes. *Microbiology (Reading, Engl)*. 2006;152(pt 12):3667–3681.
49. Al-Salih G, et al. Role of vegetation-associated protease activity in valve destruction in human infective endocarditis. *PLoS One*. 2012;7(9):e45695.
50. Wiita AP, Hsu GW, Lu CM, Esensten JH, Wells JA. Circulating proteolytic signatures of chemotherapy-induced cell death in humans discovered by N-terminal labeling. *Proc Natl Acad Sci U S A*. 2014;111(21):7594–7599.
51. Braga-Lagache S, Buchs N, Iacovache MI, Zuber B, Jackson CB, Heller M. Robust label-free, quantitative profiling of circulating plasma microparticle (MP) associated proteins. *Mol Cell Proteomics*. 2016;15(12):3640–3652.
52. Omenn GS. The Human Proteome Organization Plasma Proteome Project pilot phase: reference specimens, technology platform comparisons, and standardized data submissions and analyses. *Proteomics*. 2004;4(5):1235–1240.

53. Anavekar NS, et al. Impact of prior antiplatelet therapy on risk of embolism in infective endocarditis. *Clin Infect Dis*. 2007;44(9):1180–1186.
54. Chan KL, et al. A randomized trial of aspirin on the risk of embolic events in patients with infective endocarditis. *J Am Coll Cardiol*. 2003;42(5):775–780.
55. Lerche CJ, et al. Adjunctive dabigatran therapy improves outcome of experimental left-sided Staphylococcus aureus endocarditis. *PLoS One*. 2019;14(4):e0215333.
56. Hsu CC, et al. neutrophil extracellular traps enhance staphylococcus aureus vegetation formation through interaction with platelets in infective endocarditis. *Thromb Haemost*. 2019;119(5):786–796.
57. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*. 2017;33(18):2938–2940.
58. Sievert C. *Interactive Web-Based Data Visualization with R, plotly, and shiny*. Chapman and Hall/CRC; 2020. Accessed July 2, 2020. https://plotly-r.com.