

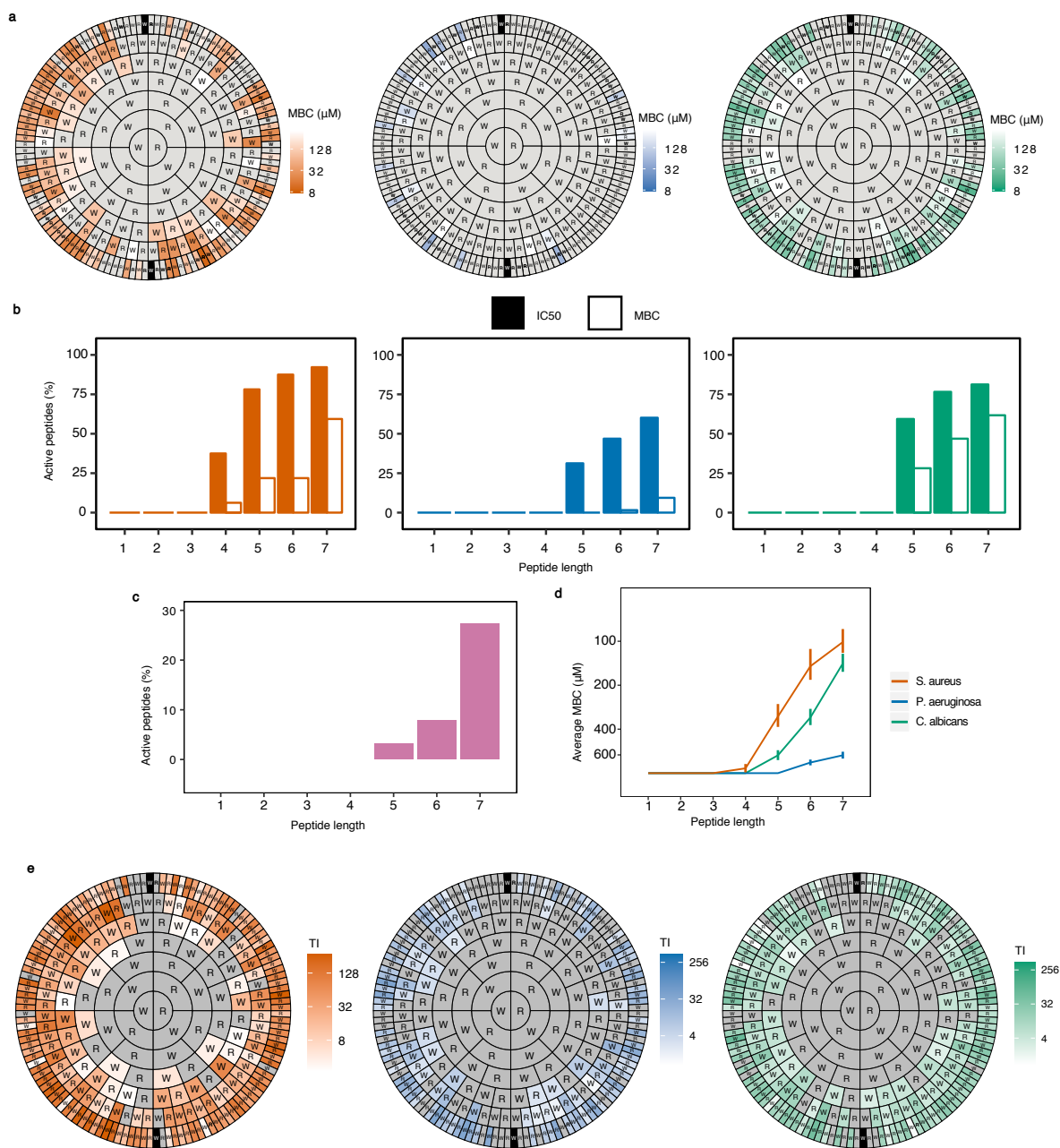
Supplementary Information for

**The Lexicon of Antimicrobial Peptides: a Complete Set of Arginine and
Tryptophan Sequences.**

Authors: Sam Clark, Thomas A Jowitt, Lynda K Harris, Christopher G Knight and Curtis B
Dobson.

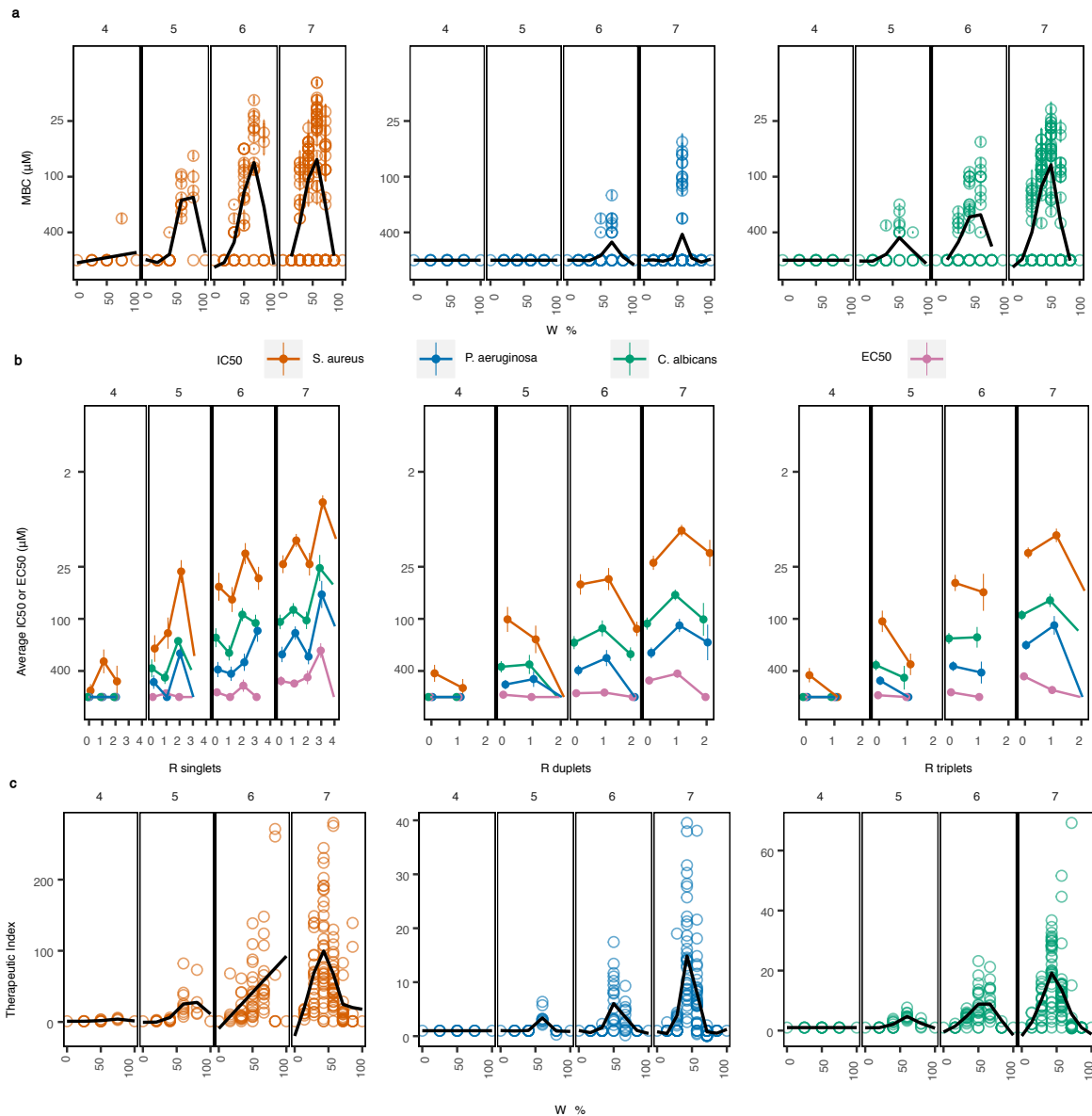
Supplementary Figures 1-5

Supplementary Tables 1 and 2



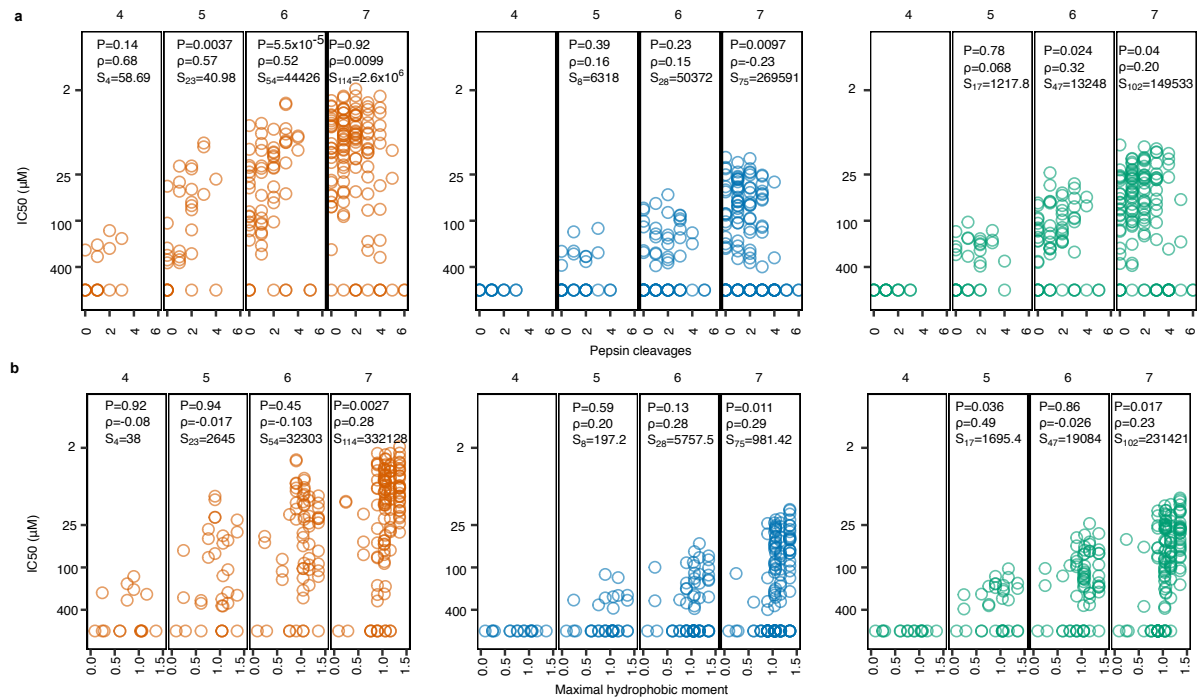
Supplementary Fig. 1: Microbicidal activity of the complete set of peptides comprised of W and R up to 7 residues long. (a) Microbicidal activities against the 3 different microorganisms studied (*S. aureus*, *P. aeruginosa* and *C. albicans*), represented using Harris-Clark diagrams. Grey sections represent peptides which did not exhibit an MBC within the range of concentrations assayed (0.8 – 400 μM). (b) Comparison of the percentage of peptides in each length subset which exhibited inhibitory or microbiocidal activity against the 3 different microorganisms within the concentration range assayed. (c) The percentage of peptides in each length subset that exhibited hemolytic activity within the concentration range assayed. (d)

Effect of peptide length on harmonic means and standard deviations of MBC. For further explanation see Fig. 1. All error bars shown are \pm s.d. ($n = 2, 4, 8, 16, 32, 64$ and 128 peptides for lengths $1 - 7$ respectively). (e) Therapeutic indices against the three microorganisms represented using Harris-Clark diagrams. Grey sections represent peptides which exhibited neither an IC_{50} nor an EC_{50} within the range of concentrations assayed ($0.8 - 400 \mu\text{M}$).

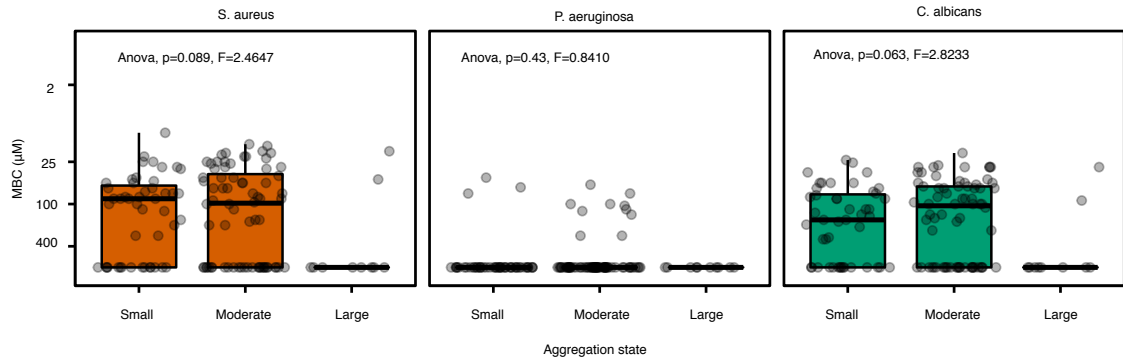


Supplementary Fig. 2: Effect of various peptide primary structural features on microbicidal activity. (a) Microbicidal activity plotted against percentage W residues within the sequence, faceted by peptide length (indicated at the top of each sub panel). The black spline through the data indicates the average activity for each peptide length. Error bars shown are \pm s.e.m. (b) Analysis of average inhibitory and hemolytic activities for peptides with

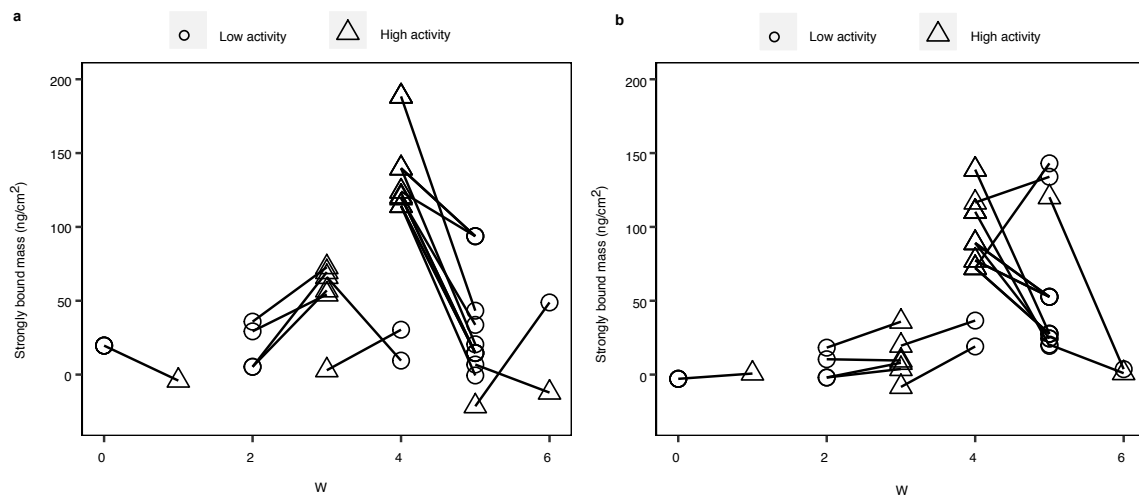
different numbers of isolated R singlets (R), duplets (RR), and triplets (RRR), faceted by peptide length (indicated at the top of each sub panel). Error bars shown are +/- s.e.m. For further explanation see Fig. 2. (c) Therapeutic index plotted against percentage W residues within the sequence, faceted by peptide length (indicated at the top of each sub panel). The black spline through the data indicates the average therapeutic index for each peptide length. Error bars shown are +/- s.e.m. (n=4).



Supplementary Fig. 3: *In silico* analysis of effect of number of pepsin cleavage sites or hydrophobic moment on antimicrobial activity. (a) Inhibitory activity plotted against number of potential pepsin cleavage sites or against (b) hydrophobic moment, faceted by peptide length. Spearman's rank correlation coefficients were calculated for each faceted dataset for peptides which exhibited IC_{50} within the concentration range assayed (0.8 – 400 μM) and are shown in each panel.



Supplementary Fig. 4: Relationship between microbiocidal activity and peptide aggregation. (a) Comparisons of microbiocidal activity and aggregation state for the three organisms assayed. Aggregation state was assessed in stock solutions (800 μM) using DLS, indicating the three size categories identified (featuring small, moderate or large aggregates). Error bars shown are \pm s.e.m. For further explanation see Fig. 3.



Supplementary Fig. 5: Relationships between selected peptide pairs and membrane binding. Mass of peptides bound to (a) anionic and (b) neutral membranes relative to the total number of W residues in each peptide. Peptide pairs in the selection are connected by lines, with the less active antimicrobial peptide indicated by a circle and the more active antimicrobial peptide by a triangle. For further explanation see Fig. 4.

Supplementary Table 1: features selected by the Boruta algorithm as having a statistically significant relationship with each of the response indices. For each index, the features have been ordered in descending level of mean importance. The number of features selected is shown, as well as the number of iterations required to reach a decision for all features.

	IC ₅₀			MBC			EC ₅₀	OD _{vis}
	S. aureus	P. aeruginosa	C. albicans	S. aureus	P. aeruginosa	C. albicans		
mass	R%	mass	mass	mass	mass	mass	mass	W
\bar{X}_R	W%	R%	R%	sequence	W	W	W	mass
length	W	W%	length	W	W%	WW	WW	R%
W%	mass	W	W	\bar{X}_W	R%	R%	R%	W%
R%	R	length	R%	W%	σ_W	W%	W%	length
W	length	X ₅	W%	W	R	R	R	R
σ_W	sequence	σ_W	σ_W	R	length	WWRRWW		
X ₅	σ_R	R	σ_R	RRRW	σ_R	RWW		
X ₆	\bar{X}_R	σ_R	WW	WWRRRW	sequence	σ_R		
sequence	σ_W	X ₆	R	RRR	WW	length		
RWWWWW	X ₇	WW	WWR	\bar{X}_R	X ₆	X ₇		
\bar{X}_W	X ₅	\bar{X}_R	RRWW	RWR	WWRRRW	RWWR		
WWWWW	X ₆	sequence	RWRWWW	RRWR	RWW	X ₆		
R	\bar{X}_W	WWR	RRWWW	RRWWW	WWR	σ_W		
X ₄	RRWW	X ₇	RRWWR	RRWWR	\bar{X}_R	WWR		
WR	WWR	WRWWW	RW	RW	\bar{X}_W	RRW		
σ_R	RWW	\bar{X}_W	RWW	RWW	X ₇	WWRWW		
WWWWWR	WW	RRWWW	\bar{X}_W	WWW	WWW	sequence		
RW	RW	RWW	X ₅	WRRR	WRRR	RWWRW		
X ₇	RRW	WR	WWRW	RWWW	RWWW	WRRRW		
WRWWW	RWR	WWWWW	sequence	X ₅	WWW	WWW		
WWRW		WRR		RW	RW			
WWR		WRR		WRR	WRR			
WWW		RRWWW		WRR	WRR			
WW		X ₄		WRRW	WRRW			
WWRWW		RRWW		RRWW	RRWW			
WRRR				RRW	RRW			
RWW				WRWW	WRWW			
WRR				RWWW	RWWW			
				WRRRR	WRRRR			
features selected	29	21	26	21	13	30	21	6

Supplementary Table 2: Table showing the feature space used to describe peptides. Features have been split into different categories. The numbers of features in each category are shown, along with the total number of features in the space. In the symbol column A is amino acid and C is amino acid class.

Symbol	Feature category	Number of features
A%	Percentage abundance of amino acid	2
\bar{X}_A	Mean position of amino acid	2
σ_A	Positional standard deviation of amino acid	2
AA	Duplets	4
AAA	Triplets	8
X ₇	Sequence position	7
	Longer sub-sequences	240
	Other features (length, mass)	2
	Total number of features	267