Leo *et al.*: A first analysis of a bacterial collagen-binding protein with the collagen Toolkits: the promiscuous binding of YadA to collagens may explain how it interferes with host processes

SUPPLEMENTARY MATERIAL

Synthesis of cyclic GPO peptides.

The cyclo$(GPO)_n$ peptides were synthesized from linear precursors off-resin. The linear PO(*t*-Bu)$(GPO(t\text{-Bu}))_{n-1}$G-OH peptides were synthesized as above (0.1 mmol scale) except that they were synthesized as C-terminal acids using the Tentagel R Trt-Gly Fmoc resin (0.17 mmol/g, Rapp Polymere). Cleavage of the side-chain protected peptides from the resin was carried out by repeated treatment with 1% trifluoroacetic acid in dichloromethane (8 ml, 2 mins, 8 times). A 10% solution of *N*,*N*-diisopropylamine in methanol was then carefully added to the previous reaction mixture. The resin was filtered and successively washed with dichloromethane (20 ml, twice), methanol (20ml, twice) and dichloromethane (20 ml, twice). The filtrate was concentrated under reduced pressure to *ca.* 1 ml volume, after which the crude peptides were precipitated with ice-cold ether. The filtered side-chain protected crude peptides were ether-washed (twice), dissolved in 10% acetonitrile in water, lyophilized and then purified by reverse phase high performance liquid chromatography as described in (36, 56) using a linear gradient of 5-45% acetonitrile in water. The pure peptides were then lyophilized.

Cyclization of the side-chain protected peptides (35 μmol) was carried out using HATU (2-(7-Aza-1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate, 3 eq.) and HOBt (N-Hydroxybenzotriazole, 3 eq) in dimethylformamide (15 ml) containing 1% *N*,*N*-diisopropylamine, overnight. The solvent was removed under reduced pressure and the resulting residue was dissolved in dichloromethane. The organic layer was then successively washed with water, a saturated solution of sodium carbonate, a 1 M solution of potassium hydrogen sulfate and water. Removal of the dichloromethane under reduced pressure gave the crude side-chain protected cyclic peptides. Side-chain deprotection was effected by treatment with a trifluoroacetic acid, water and triisopropylsilane mixture (95:2.5:2.5 v/v), for 1 h. The reaction mixture was then concentrated under reduced pressure to *ca.* 1 ml volume, after which the crude cyclic peptides were precipitated with ice-cold ether. The filtered crude cyclic peptides were ether-washed (twice), dissolved in 5% acetonitrile in water and then lyophilized. The crude cyclic peptides were purified, as described for the linear precursors, characterized by MALDI-TOF mass spectrometry, and then lyophilized.

Supplementary Table 1
Sequences of Toolkit peptides

Toolkit II		Toolkit III	
Peptide	Sequence[a]	Peptide	Sequence[a]
II-1	GPMGPMGPRGPOGPAGAOGPQGFQGNO	III-1	GLAGYOGPAGPOGPOGPOGTSGHOGSO
II-2	GPQGFQGNOGEOGEOGVSGPMGPRGPO	III-2	GTSGHOGSOGSOGYQGPOGEOGQAGPS
II-3	GPMGPRGPOGPOGKOGDDGEAGKOGKA	III-3	GEOGQAGPSGPOGPOGAIGPSGPAGKD
II-4	GEAGKOGKAGERGPOGPQGARGFOGTO	III-4	GPSGPAGKDGESGROGROGERGLOGPO
II-5	GARGFOGTOGLOGVKGHRGYOGLDGAK	III-5	GERGLOGPOGIKGPAGIOGFOGMKGHR
II-6	GYOGLDGAKGEAGAOGVKGESGSOGEN	III-6	GFOGMKGHRGFDGRNGEKGETGAOGLK
II-7	GESGSOGENGSOGPMGPRGLOGERGRT	III-7	GETGAOGLKGENGLOGENGAOGPMGPR
II-8	GLOGERGRTGPAGAAGARGNDGQOGPA	III-8	GAOGPMGPRGAOGERGROGLOGAAGAR
II-9	GNDGQOGPAGPOGPVGPAGGOGFOGAO	III-9	GLOGAAGARGNDGARGSDGQOGPOGPO
II-10	GGOGFOGAOGAKGEAGPTGARGPEGAQ	III-10	GQOGPOGPOGTAGFOGSOGAKGEVGPA
II-11	GARGPEGAQGPRGEOGTOGSOGPAGAS	III-11	GAKGEVGPAGSOGSNGAOGQRGEOGPQ
II-12	GSOGPAGASGNOGTDGIOGAKGSAGAO	III-12	GQRGEOGPQGHAGAQGPOGPOGINGSO
II-13	GAKGSAGAOGIAGAOGFOGPRGPOGPQ	III-13	GPOGINGSOGGKGEMGPAGIOGAOGLM
II-14	GPRGPOGPQGATGPLGPKGQTGEOGIA	III-14	GIOGAOGLMGARGPOGPAGANGAOGLR
II-15	GQTGEOGIAGFKGEQGPKGEOGPAGPQ	III-15	GANGAOGLRGGAGEOGKNGAKGEOGPR
II-16	GEOGPAGPQGAOGPAGEEGKRGARGEO	III-16	GAKGEOGPRGERGEAGIOGVOGAKGED
II-17	GKRGARGEOGGVGPIGPOGERGAOGNR	III-17	GVOGAKGEDGKDGSOGEOGANGLOGAA
II-18	GERGAOGNRGFOGQDGLAGPKGAOGER	III-18	GANGLOGAAGERGAOGFRGPAGPNGIO
II-19	GPKGAOGERGPSGLAGPKGANGDOGRO	III-19	GPAGPNGIOGEKGPAGERGAOGPAGPR
II-20	GANGDOGROGEOGLOGARGLTGROGDA	III-20	GAOGPAGPRGAAGEOGRDGVOGGOGMR
II-21	GLTGROGDAGPQGKVGPSGAOGEDGRO	III-21	GVOGGOGMRGMOGSOGGOGSDGKOGPO
II-22	GAOGEDGROGPOGPQGARGQOGVMGFO	III-22	GSDGKOGPOGSQGESGROGPOGPSGPR
II-23	GQOGVMGFOGPKGANGEOGKAGEKGLO	III-23	GPOGPSGPRGQOGVMGFOGPKGNDGAO
II-24	GKAGEKGLOGAOGLRGLOGKDGETGAA	III-24	GPKGNDGAOGKNGERGGOGGOGPQGPO
II-25	GKDGETGAAGPOGPAGPAGERGEQGAO	III-25	GGOGPQGPOGKNGETGPQGPOGPTGPG
II-26	GERGEQGAOGPSGFQGLOGPOGPOGEG	III-26	GPOGPTGPGGDKGDTGPOGPQGLQGLO
II-27	GPOGPOGEGGKOGDQGVOGEAGAOGLV	III-27	GPQGLQGLOGTGGPOGENGKOGEOGPK
II-28	GEAGAOGLVGPRGERGFOGERGSOGAQ	III-28	GKOGEOGPKGDAGAOGAOGGKGDAGAO
II-29	GERGSOGAQGLQGPRGLOGTOGTDGPK	III-29	GGKGDAGAOGERGPOGLAGAOGLRGGA
II-30	GTOGTDGPKGASGPAGPOGAQGPOGLQ	III-30	GAOGLRGGAGPOGPEGGKGAAGPOGPO
II-31	GAQGPOGLQGMOGERGAAGIAGPKGDR	III-31	GAAGPOGPOGAAGTOGLQGMOGERGGL
II-32	GIAGPKGDRGDVGEKGPEGAOGKDGGR	III-32	GMOGERGGLGSOGPKGDKGEOGGOGAD
II-33	GAOGKDGGRGLTGPIGPOGPAGANGEK	III-33	GEOGGOGADGVOGKDGPRGPTGPIGPO

II-34	GPAGANGEKGEVGPOGPAGSAGARGAO	III-34	GPTGPIGPOGPAGQOGDKGEGGAOGLO
II-35	GSAGARGAOGERGETGPOGPAGFAGPO	III-35	GEGGAOGLOGIAGPRGSOGERGETGPO
II-36	GPAGFAGPOGADGQOGAKGEQGEAGQK	III-36	GERGETGPOGPAGFOGAOGQNGEOGGK
II-37	GEQGEAGQKGEAGAOGPQGPSGAOGPQ	III-37	GQNGEOGGKGERGAOGEKGEGGPOGVA
II-38	GPSGAOGPQGPTGVTGPKGARGAQGPO	III-38	GEGGPOGVAGPOGGSGPAGPOGPQGVK
II-39	GARGAQGPOGATGFOGAAGRVGPOGSN	III-39	GPOGPQGVKGERGSOGGOGAAGFOGAR
II-40	GRVGPOGSNGNOGPOGPOGPSGKDGPK	III-40	GAAGFOGARGLOGPOGSNGNOGPOGPS
II-41	GPSGKDGPKGARGDSGPOGRAGEOGLQ	III-41	GNOGPOGPSGSOGKDGPOGPAGNTGAO
II-42	GRAGEOGLQGPAGPOGEKGEOGDDGPS	III-42	GPAGNTGAOGSOGVSGPKGDAGQOGEK
II-43	GEOGDDGPSGAEGPOGPQGLAGQRGIV	III-43	GDAGQOGEKGSOGAQGPOGAOGPLGIA
II-44	GLAGQRGIVGLOGQRGERGFOGLOGPS	III-44	GAOGPLGIAGITGARGLAGPOGMOGPR
II-45	GFOGLOGPSGEOGKQGAOGASGDRGPO	III-45	GPOGMOGPRGSOGPQGVKGESGKOGAN
II-46	GASGDRGPOGPVGPOGLTGPAGEOGRE	III-46	GESGKOGANGLSGERGPOGPQGLOGLA
II-47	GPAGEOGREGSOGADGPOGRDGAAGVK	III-47	GPQGLOGLAGTAGEOGRDGNOGSDGLO
II-48	GRDGAAGVKGDRGETGAVGAOGAOGPO	III-48	GNOGSDGLOGRDGSOGGKGDRGENGSO
II-49	GAOGAOGPOGSOGPAGPTGKQGDRGEA	III-49	GDRGENGSOGAOGAOGHOGPOGPVGPA
II-50	GKQGDRGEAGAQGPMGPSGPAGARGIQ	III-50	GPOGPVGPAGKSGDRGESGPAGPAGAO
II-51	GPAGARGIQGPQGPRGDKGEAGEOGER	III-51	GPAGPAGAOGPAGSRGAOGPQGPRGDK
II-52	GEAGEOGERGLKGHRGFTGLQGLOGPO	III-52	GPQGPRGDKGETGERGAAGIKGHRGFO
II-53	GLQGLOGPOGPSGDQGASGPAGPSGPR	III-53	GIKGHRGFOGNOGAOGSOGPAGQQGAI
II-54	GPAGPSGPRGPOGPVGPSGKDGANGIO	III-54	GPAGQQGAIGSOGPAGPRGPVGPSGPO
II-55	GKDGANGIOGPIGPOGPRGRSGETGPA	III-55	GPVGPSGPOGKDGTSGHOGPIGPOGPR
II-56	GPRGRSGETGPAGPOGNOGPOGPOGPO	III-56	GPIGPOGPRGNRGERGSEGSOGHOGQO
		III-57	GERGSEGSOGHOGQOGPOGPOGAOGPC

[a]Flanking GPC-$(GPP)_5$ sequences are not shown

Supplementary Table 2

Melting temperatures of GPO6 peptides as measured by polarimetry.

Peptide	T_m
GPO6a	73.3°C
GPO6b	74.5°C
GPO6c	74.1°C
GPO6d	73.8°C
GPO6e	73.0°C
GPO6f	73.6°C
GPO6g	73.9°C
GPO6h	74.0°C
GPO6i	74.1°C
GPO6j	74.3°C
GKO-GPO1	50.9°C
GKO-GPO3	63.8°C
GKO-GPO4	64.6°C
GKO-GPO6	59.4°C

Supplementary Table 3
ERGET-like motifs in Toolkit peptides

Peptide	Motif	Binding class	Peptide	Motif	Binding class
II-7	ERGRT	Non-binder	III-4	KDGES	Medium binder
II-8	ERGRT	Non-binder	III-11	QRGEO	Low binder
II-24	KDGET	Non-binder	III-12	QRGEO	High binder
II-25	KDGET	Low binder	III-15	AKGEO	Low binder
II-25	ERGEQ	Low binder	III-16	ERGEA	Non-binder
II-26	ERGEQ	High binder	III-16	AKGED	Non-binder
II-32	DRGDV	Non-binder	III-17	AKGED	Non-binder
II-34	EKGEV	Non-binder	III-25	KNGET	Medium binder
II-35	ERGET	High binder	III-26	DKGDT	High binder
II-36	EQGEA	Non-binder	III-32	DKGEO	Non-binder
II-42	EKGEO	Low binder	III-34	DKGEG	Medium binder
II-44	QRGER	Medium binder	III-35	ERGET	Non-binder
II-48	DRGET	Medium binder	III-36	ERGET	Low binder
II-49	DRGEA	Medium binder	III-37	EKGEG	Non-binder
II-50	DRGEA	Non-binder	III-45	VKGES	Low binder
II-51	DKGEA	Non-binder	III-48	DRGEN	Non-binder
II-56	RSGET	High binder	III-49	DRGEN	Medium binder
			III-56	NRGER	Low binder
			III-57	ERGSE	Medium binder

Supplementary Table 4

Correlations of amino acid content of Toolkit peptides to YadA binding[a].

Amino Acid	Toolkit II (N=56)			Toolkit III (N=55)			Toolkit II & III (N=111)		
	#[b]	Correlation	*p*-value	#[b]	Correlation	*p*-value	#[b]	Correlation	*p*-value
Alanine	159	-0.195	0.150	133	-0.006	0.963	292	-0.105	0.272
Arginine	84	-0.063	0.641	63	-0.177	0.195	147	-0.117	0.220
Asparagine	20	-0.090	0.509	33	0.010	0.943	53	-0.029	0.764
Aspartic acid	44	-0.262	0.051	34	-0.132	0.336	78	-0.201	0.034
Glutamic acid	84	-0.153	0.209	64	-0.409	0.002	148	-0.268	0.004
Glutamine	58	0.0952	0.444	38	0.067	0.629	96	0.074	0.421
Glycine	511	0.170	0.426	524	-0.211	0.121	1035	-0.063	0.517
Histidine	2	0.057	0.676	9	0.164	0.231	11	0.119	0.212
Hydroxyproline	168	0.559	<0.0001	210	0.320	0.017	378	0.420	<0.0001
Isoleucine	15	-0.231	0.087	21	0.038	0.781	36	-0.085	0.376
Leucine	36	-0.022	0.871	36	0.080	0.563	72	0.049	0.604
Lysine	56	-0.523	<0.0001	50	-0.495	0.0001	106	-0.511	<0.0001
Methionine	9	-0.140	0.303	14	0.017	0.904	23	0.0771	0.421
Phenylalanine	19	0.420	0.001	8	0.075	0.588	27	0.270	0.004
Proline	150	0.217	0.108	152	0.306	0.023	302	0.258	0.006
Serine	43	-0.049	0.720	56	0.061	0.657	99	0.010	0.915
Threonine	29	-0.045	0.745	19	0.236	0.083	48	0.074	0.443
Tyrosine	2	-0.126	0.352	2	0.277	0.040	4	0.063	0.510
Valine	23	0.104	0.444	18	-0.164	0.232	41	-0.016	0.870

[a]Correlations at higher than the 0.05 significance level are underlined
[b]The total number of this type of residue in the Toolkit peptides

Supplementary Table 5

Occurences of common X-X' dipeptides in the 15 highest and lowest-binding Toolkit peptides[a]

Dipeptide	Occurrence in high binders	Occurrence in low binders	Expected occurrence based on Toolkit frequencies	Expected occurrence based on frequencies in collagens[b]
AA	3	3	2.38	1.215
AK	0	2	1.39	1.62
AN	0	4	1.39	0.405
AO	7	11	8.52	4.59
AQ	1	1	1.39	0.54
AR	5	1	2.97	1.485
DK	1	2	0.99	1.35
DR	1	3	1.19	3.645
EA	1	3	1.78	1.62
EK	0	6	1.59	3.375
EN	0	1	0.99	0.405
EO	2	5	4.56	3.78
ER	5	5	4.96	3.645
ES	1	0	0.99	0.405
ET	3	1	1.59	0.81
FO	4	1	2.78	3.375
GO	1	1	1.39	0.405
IA	0	2	0.99	0.54
IO	0	2	1.19	2.025
KD	1	4	1.78	0.675
KO	0	0	1.19	1.89
LA	2	1	1.39	0.675
LO	8	4	3.77	7.425
LQ	2	1	1.19	0.945
NO	3	1	1.39	0.945
PA	7	8	7.53	4.59
PI	0	1	0.99	1.215
PK	0	4	2.58	3.645
PM	2	0	1.19	0.54
PO	25	8	14.47	14.175
PQ	4	3	3.57	3.375
PR	2	3	3.96	3.51
PS	4	3	3.17	1.89
PT	1	0	0.99	1.08
PV	1	1	0.99	1.755
QO	3	1	1.59	1.485
RO	3	0	1.39	1.485
SO	2	4	4.96	3.105
VK	1	1	0.99	0.675

[a]15 highest-binding peptides: II-1, II-9; II-26, II-56, II-35, II-48, II-22, II-44, II-52, III-1, III-9, III-40, III-12, III-26, III-4

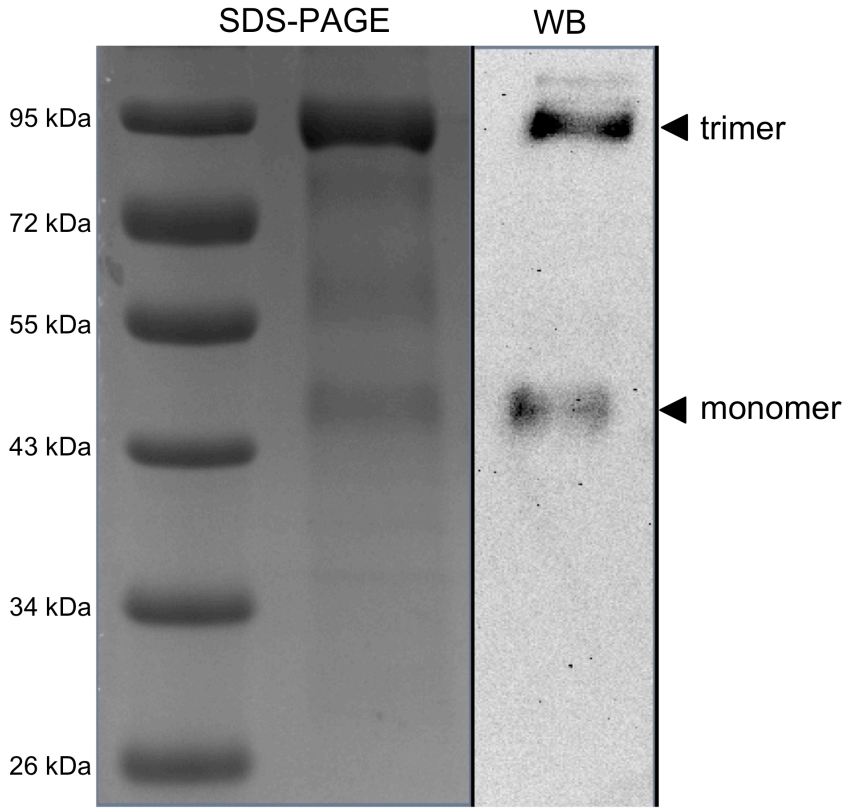
15 lowest-binding peptides: II-24, II-23, II-31, II-47, II-34, II-33, II-54, II-37, III-32, III-37, III-29, III-32, III-48, III-51

[b]expected frequencies from (42)

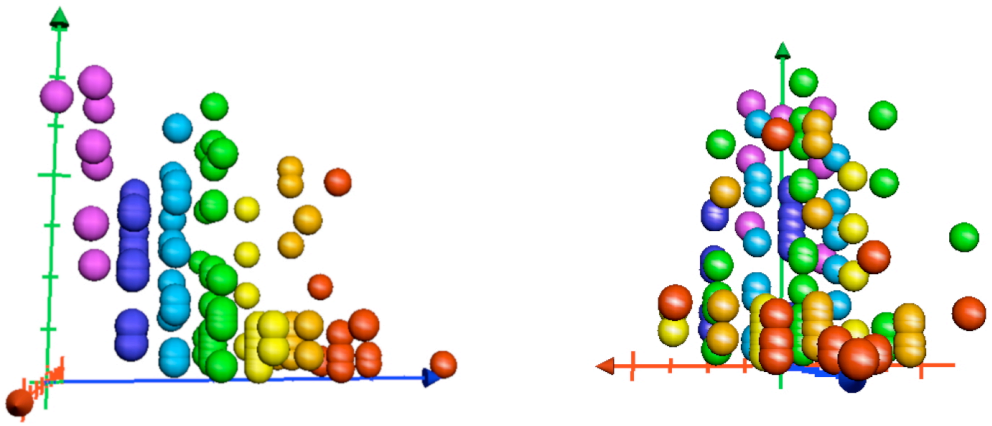
Supplementary Table 6

Significant correlations for dipeptides in Toolkit peptides with YadA binding (N=111).

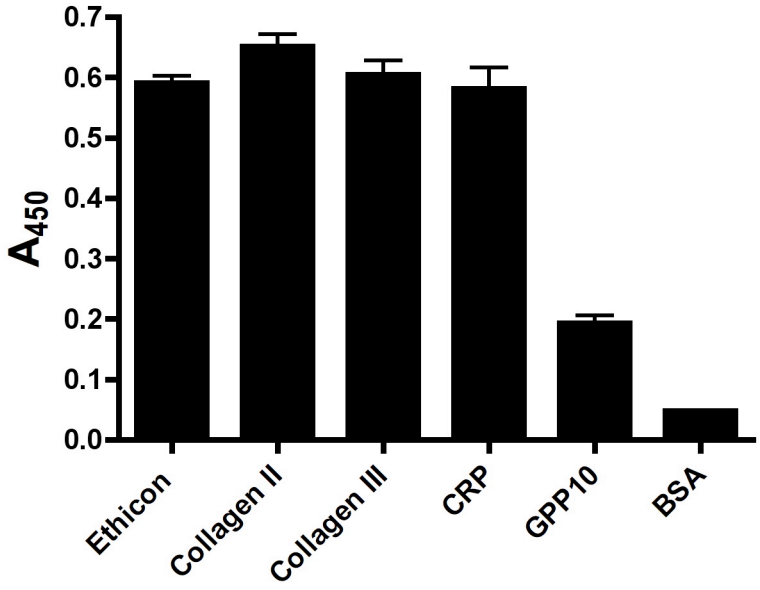
Position	Dipeptide	Correlation	*p*-value
G-X			
	GF	0.200	0.034
	GP	0.272	0.003
	GE	-0.286	0.002
X- X'			
	FQ	0.303	0.001
	EK	-0.290	0.002
	PK	-0.30203	0.001
	PO	0.47678	9.3×10^{-8}
X'-G			
	KG	-0.39248	1.71×10^{-5}
	OG	0.27531	0.003164



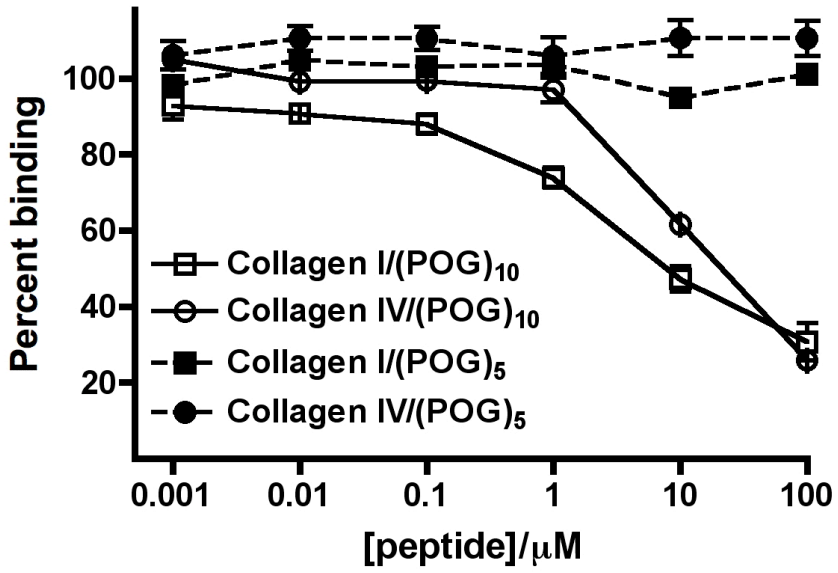
Supplementary Figure 1. $YadA_{24\text{-}378}$ is trimeric. 5 μg of purified $YadA_{24\text{-}378}$ was heated for 10 minutes in non-reducing sample buffer and subjected to SDS-PAGE (left panel). Most of the protein is still trimeric, which migrates somewhat anomalously. The identity of the protein was confirmed by Western blot using the YadA-specific antibody 3G12 (right panel).



Supplementary Figure 2. 3-dimensional scatter plot of the effect of the number of charged residues and net charge of the Toolkit peptides on binding response. The green axis shows the binding response (A_{450}), the blue axis the number of charged residues and the red axis the net charge of the peptide. To aid interpretation, the points are coloured according to the number of charged residues in the peptide: purple 0-1 charges, dark blue 2 charges, light blue 3 charges, green 4 charges, yellow 5 charges, orange 6 charges, red 7-10 charges. When comparing points with the same number of charges, on average those with a net charge of 0±1 give a higher binding response than those with a higher absolute charge. The magnitude of absolute charge also decreases the binding response, *i.e.* peptides with a smaller absolute charge tend to bind YadA better than those with a larger absolute charge. This suggests that neutralising charges can compensate to some degree for the number of charged residues, which have an inhibitory effect on YadA binding.

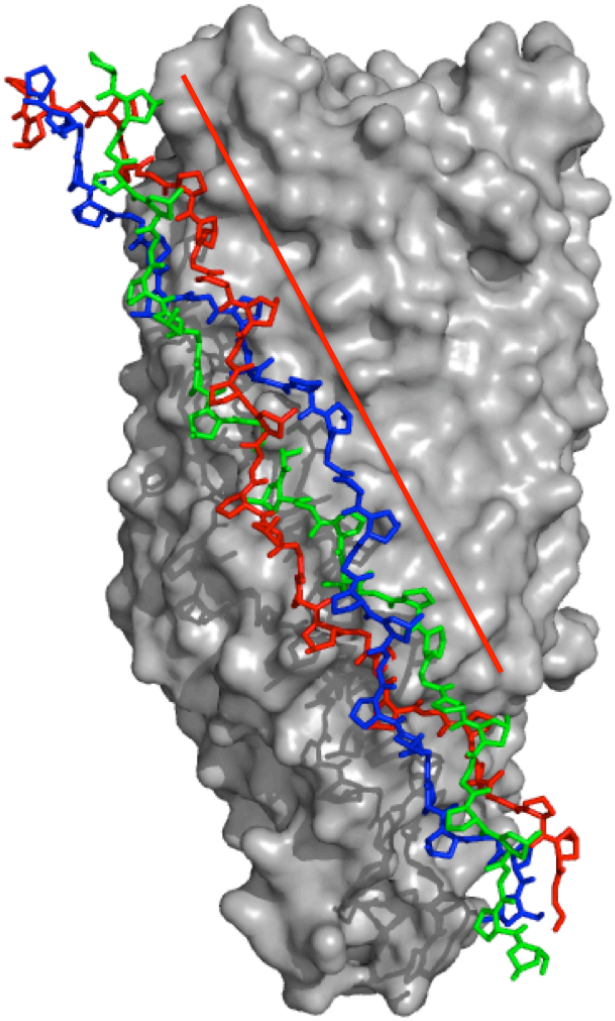


Supplementary Figure 3. Binding of YadA to collagens and CRP by SPBA. Ethicon collagen is a fibrous form of collagen type I. Collagens type II and type III are monomeric. CRP is a cross-linked GPO-containing peptide, and GPP10 is a control peptide. Binding to BSA shows background levels. The columns show the mean absorbance at 450 nm from three replicate wells; error bars denote standard errors of mean. YadA bound as strongly to this as to monomeric collagens type II and type III. Additionally, YadA bound to CRP at a similar level. This was expected, as CRP is composed of cross-linked GCO-$(GPO)_{10}$-GCOG peptides, and YadA is known to bind strongly to the peptide $(POG)_{10}$ (30).



Supplementary Figure 4. $(POG)_{10}$ blocks the binding of YadA to collagen type I and type IV. Blocking experiments were carried out using SPBA. Wells of Immulon 2HB plates were coated with either collagen type I or collagen type IV as described for SPBA in Materials and Methods. YadA was diluted to 10 μg/ml in adhesion buffer to which $(POG)_{10}$ or $(POG)_5$ were added at concentrations between 0 and 100 μM per trimer. These solutions were incubated at RT for 30 minutes before use. After blocking the wells with BSA as above, we added 100 μl of the YadA-peptide solutions and incubated for 1 hour. The procedure after this was as for SPBA above. For analysis, we took wells incubated without POG peptides as 100% of binding, and the background level represented by BSA as 0%. Binding levels were calculated as follows: $P = [(S_{col}\text{-}S_{BSA}) \text{ x } (S_{POG}\text{-}S_{BSA})] \text{ x } 100\%$, where P is the percentage of binding, and S_{col}, S_{POG} and S_{BSA} the signals given by collagen without added peptide, the signal given at a given peptide concentration and the signal from BSA, respectively.

The results for collagen type I (squares) and collagen type IV (circles) are shown above. $(POG)_{10}$ (open symbols and continuous curves) blocked the binding of YadA to collagen type I at concentrations of 1 μM or higher (Figure 7). $(POG)_{10}$ also blocked binding to collagen type IV, yielding a similar inhibition curve as for collagen type I. $(POG)_5$ (filled symbols and dashed curves), which is not triple-helical, did not block the interaction to either collagen at any of the assayed concentrations. Though these results do not exclude the possibility that YadA binds specifically to the 7sL fragment, they do suggest that YadA binds to triple-helical regions in collagen type IV as well.



Supplementary Figure 5. Model of $(POG)_{10}$ binding to YadA. The three chains of the peptide are shown in line representation and are coloured differently. The line in red demonstrates the length of the 7 POG repeats that interact with YadA. The model was produced by docking the peptide onto the YadA surface based on mutagenesis data as described in (39).