Supplementary Information for

Elucidating structure-performance relationships in whole-cell cooperative enzyme catalysis

Mason R. Smith,[‡a] *Hui Gao,*[‡ab] *Ponnandy Prabhu,*[a] *Luke F. Bugada,*[ac] *Cori Roth,*[a] *Deepika Mutukuri,*[a] *Christine M. Yee,*[a] *Lester Lee,*[a] *Robert M. Ziff,*[a] *Jung-Kul Lee,**[b] *Fei Wen**[ac]

[a]Department of Chemical Engineering, University of Michigan, Ann Arbor, Michigan 48109, United States

[b]Department of Chemical Engineering, Konkuk University, Seoul 05029, Republic of Korea

[c]Catalysis Science and Technology Institute, University of Michigan, Ann Arbor, Michigan 48109, United States

Contents

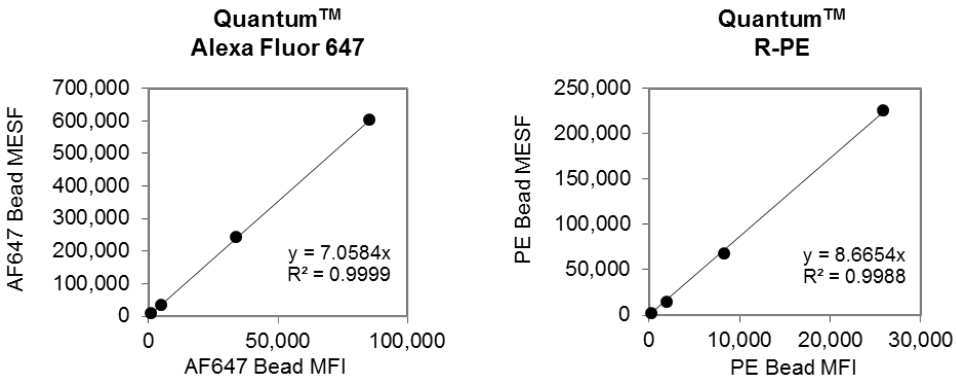
Supplementary Methods

Recombinant plasmid construction

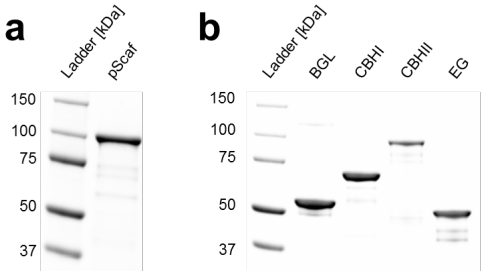
Genes encoding one (aScaf1), two (aScaf2), or three (aScaf3) type II cohesins from *C. thermocellum* OlpB scaffoldin were amplified by PCR. aScaf1 and aScaf2 were isolated and amplified from the *C. thermocellum* genome using the primer pairs aScaf1-For/aScaf1-Rev and aScaf1-For/aScaf2-Rev, respectively. aScaf3 was designed by adding an additional aScaf1-cohesin unit to the C-terminus of aScaf2 via PCR. The amplified fragments were digested with restriction enzymes (NheI/ApaI or NheI/BamHI) and ligated into similarly digested pYD1 plasmids, which include an N-terminal a-agglutinin subunit Aga2 protein and a C-terminal V5-tag. pYD1-Aga2-V5 was created by ligating pYD1 and the fragment of Aga2-V5 after NheI/PmeI-digestion, and transforming the ligation product into MachI. Positive clones of pYD1-aScaf1, pYD1-aScaf2, pYD1-aScaf3 and pYD1-Aga2-V5 were verified by DNA sequencing. PCR fragments CipA-CBM, ScaB, CipC, CbpA and type II dockerin were obtained by using primer pairs and templates as shown in Table S2. The PCR products and BamHI/XbaI-digested pRS415 were used to construct pRS415-pScaf by DNA assembly method described elsewhere.[1] pRS415-pScaf was purified from *S. cerevisiae* and then transformed into MachI, recovered, and confirmed by DNA sequencing. The fragment pScaf was obtained by performing PCR using primer pairs pScaf-For/pScaf-Rev with pRS415-pScaf as template. The plasmids used for *E. coli* expression were constructed based on pET28a. The pET28a-pScaf was created by ligating linear pET28a and pScaf after NheI/XhoI-digestion and transforming the ligation product into MachI cells. Fusion construct BGL encoding BglA and dockerin ExgS from *C. cellulovorans*, and fusion construct EG encoding CelA catalytic region from *C. thermocellum* and dockerin ScaA from *R. flavefaciens* were generated by overlap extension PCR[2] with

respective primers (Table S2). Fragments CBHI (CelF) and CBHII (CelK) with native dockerin were PCR-amplified from *C. thermocellum* and *C. cellulolyticum*, respectively. The fragments BGL, EG, CBHI and CBHII were cloned into pET28a through digestion and ligation and the cloning result was confirmed by DNA sequencing. All restriction enzymes were obtained from New England BioLabs (Ipswich, MA) and all primers were purchased from Integrated DNA Technologies (Coralville, IA).

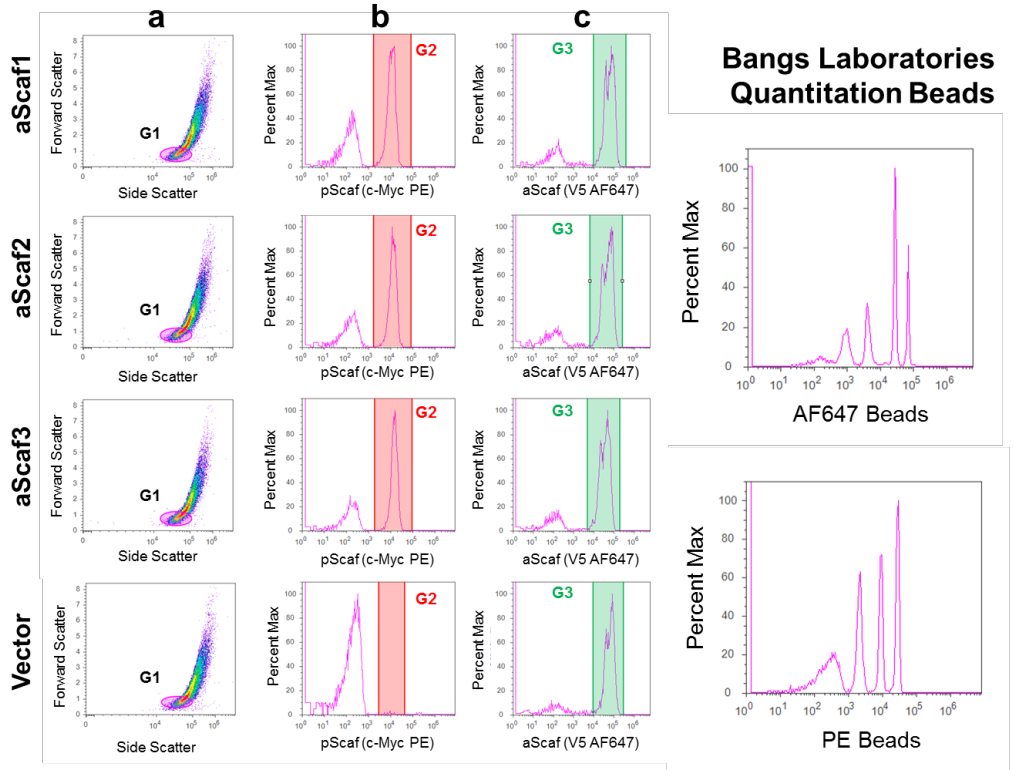
Supplementary Figures



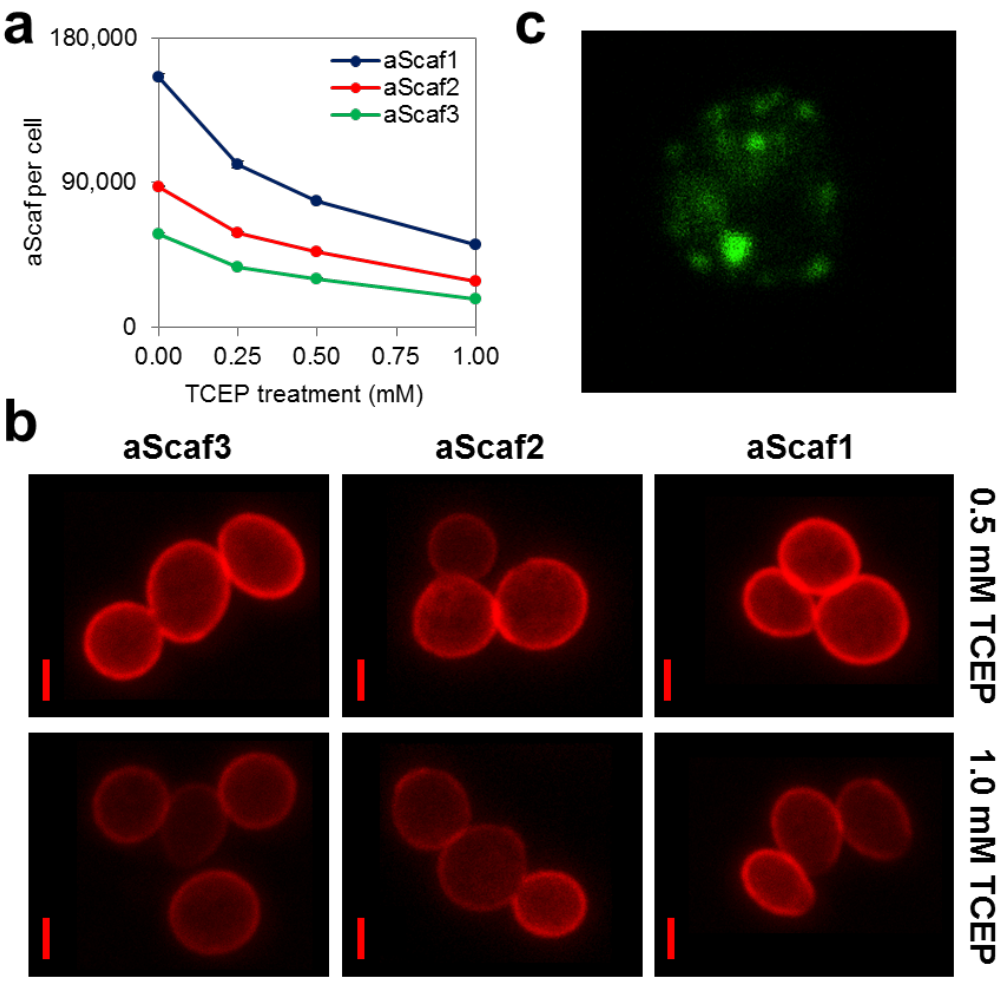
Supplementary Figure 1. Representative quantitative flow cytometry standard curves. Standard curves were created separately for each quantification experiment using Quantum™ Alexa Fluor 647 MESF and Quantum™ R-PE MESF fluorescence quantitation beads. The fluorescence quantification beads were coated with four distinct numbers of molecules of equivalent soluble fluorophore (MESF), which were provided by the manufacturer. Quantitation beads were analyzed by flow cytometry and the median fluorescence intensity (MFI) at each level of fluorophore (MESF) was plotted. These standard curves were then used to determine the number of aScaf, pScaf, and enzymes assembled on the yeast cell surface, based on their respective MFI.



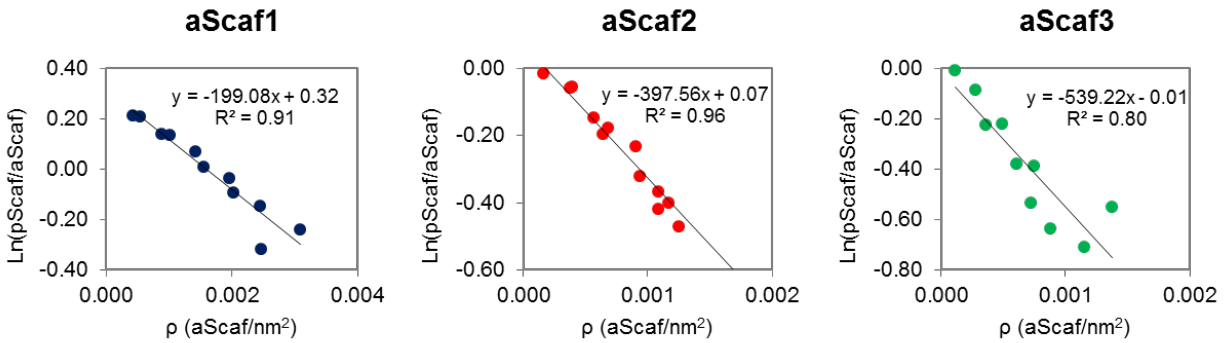
Supplementary Figure 2. SDS-PAGE analysis of **a.** pScaf and **b.** enzymes purified from *E. coli*. Each protein was purified from *E. coli* cell lysate and analyzed after boiling under reducing conditions.



Supplementary Figure 3. Flow cytometry gating scheme for quantifying aScaf-pScaf assembly. Rows correspond to different yeast constructs. Column **a** shows gating (G1) based on forward-scatter and side-scatter, which serves as parent gate for all yeast-cell analysis. Column **b** shows histogram gating (G2) on pScaf/c-Myc-positive cells. Column **c** shows histogram gating based on aScaf/V5-positive cells. Far right column shows histograms of the fluorescence intensity of the quantitation beads used in this experiment. Quantitation bead histograms were used to create standard curves as shown in Figure S1. The MFI of gates G2 and G3 were used to quantify the median number of pScaf and aScaf, respectively, displayed on each cell surface.



Supplementary Figure 4: TCEP treatment of yeast displaying each aScaf construct. **a.** aScaf display level plotted as a function of TCEP treatment (mM) for aScaf1, aScaf2, and aScaf3**. b.** Fluorescence microscopy demonstrating that surface displayed aScaf remains randomly distributed following 0.5 mM and 1.0 mM TCEP treatment. Scale bar corresponds to 2 μm. **c.** Fluorescence microscopy imaging of non-randomly distributed protein displayed on the yeast cell surface shown for comparison.

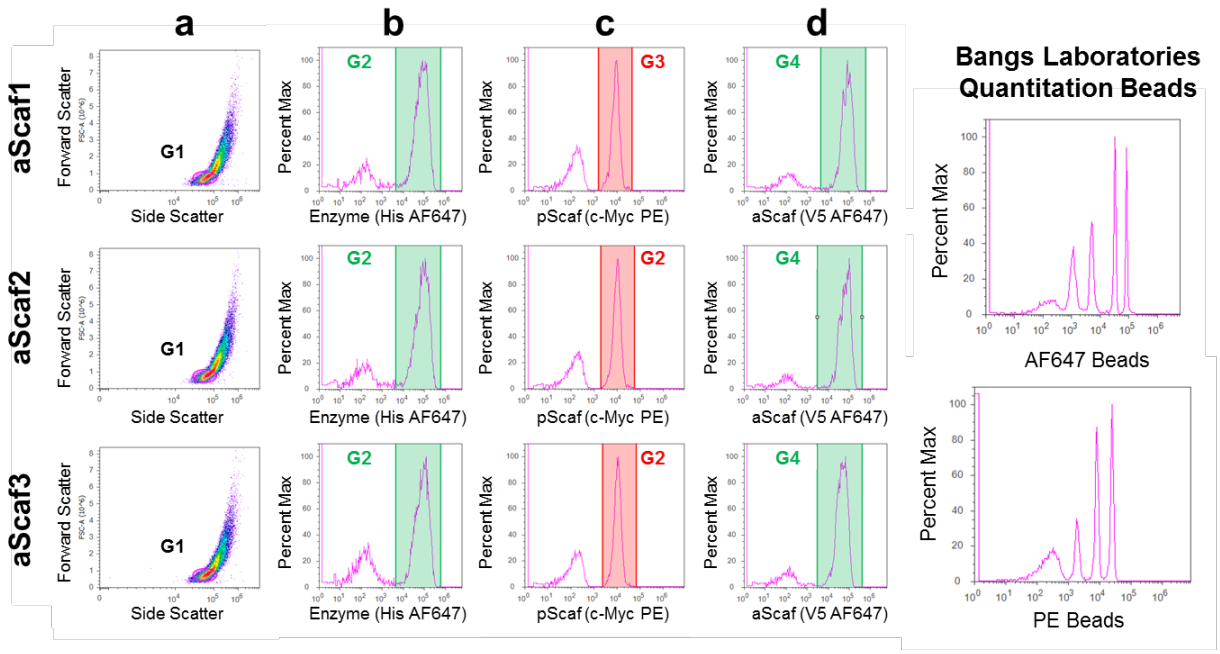


m	-199.08
S.E. *m*	20.47
R²	0.91
A	0.32
S.E. A	0.04
S.E. *y*	0.06
DF	9.00
t-value	2.26
Delta *m*	46.30
Delta *A*	0.08
Delta *y*	0.04

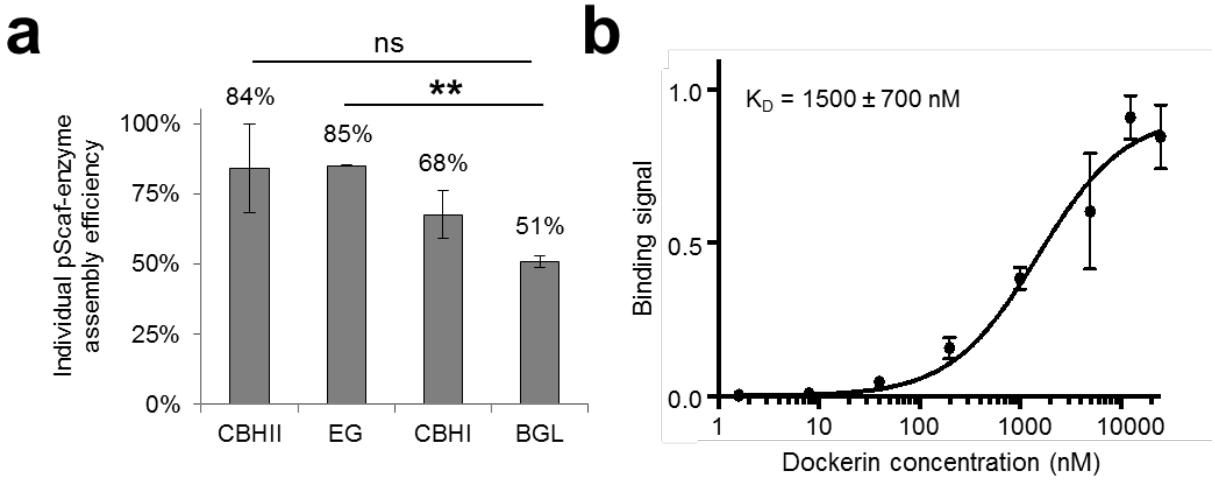
m	-397.56
S.E. *m*	22.81
R²	0.96
A	0.07
S.E. A	0.02
S.E. *y*	0.04
DF	12.00
t-value	2.18
Delta *m*	49.69
Delta *A*	0.05
Delta *y*	0.02

m	-539.22
S.E. *m*	96.16
R²	0.80
A	-0.01
S.E. A	0.07
S.E. *y*	0.11
DF	8.00
t-value	2.31
Delta *m*	221.74
Delta *A*	0.17
Delta *y*	0.08

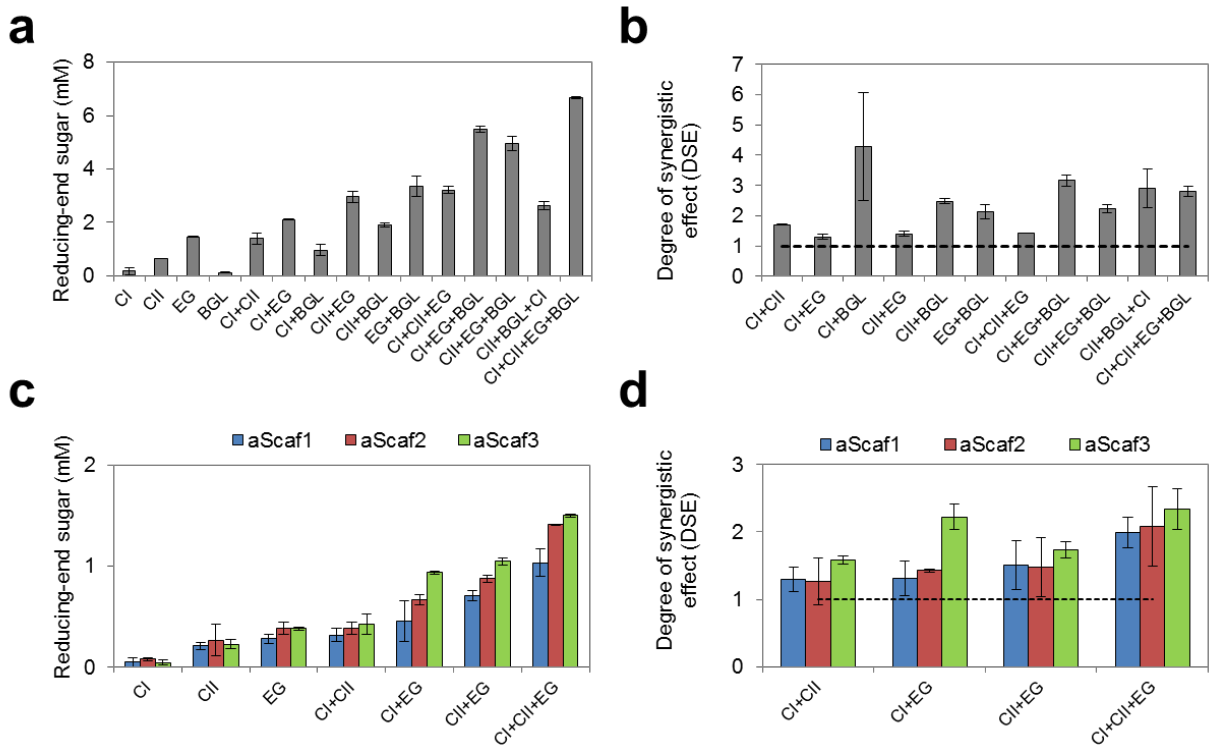
Supplementary Figure 5. Linear regression of ln-transformed aScaf-pScaf assembly data where $y = \ln\left(\frac{pScaf}{\beta aScaf}\right)$ and $x = \rho$, assuming Poisson-process distribution of surface displayed aScaf. The slope, m, corresponds to the critical distance for each aScaf construct and the intercept, A, corresponds to a correction factor that satisfies the condition that the pScaf per cell is zero when the aScaf display level is zero (See methods for details). S.E. represents the standard error of each respective regression parameter and DF represents the degrees of freedom for each dataset. R^2 represents the coefficient of determination of the linear regression with respect to the experimental data. The t-value represents the t-value of the regression coefficient with respect to the experimental data. Delta m, Delta A, and Delta y represent the margin of error of each respective parameter determined by the linear regression. Delta m and Delta A were calculated as the product of the t-value and the standard error of each respective parameter and Delta y was calculated as the product of the t-value and the standard error of y divided by the (2+DF).



Supplementary Figure 6. Flow cytometry gating scheme for quantifying overall mSEA assembly. Rows correspond to different yeast constructs. Column **a** shows gating (G1) based on forward-scatter and side-scatter, which serves as parent gate for all yeast-cell analysis. Column **b** shows histogram gating (G2) on enzyme/His-positive cells. Column **c** shows histogram gating (G3) based on pScaf/c-Myc-positive cells. Column **d** shows histogram gating (G4) based on aScaf/V5-positive cells. Far right column shows histograms of the fluorescence intensity of the quantitation beads used in this experiment. Quantitation bead histograms were used to create standard curves as shown in Figure S1. The MFI of gates G2, G3, and G4 were used to quantify the median number of enzymes, pScafs, and aScafs, respectively, displayed on each cell surface.

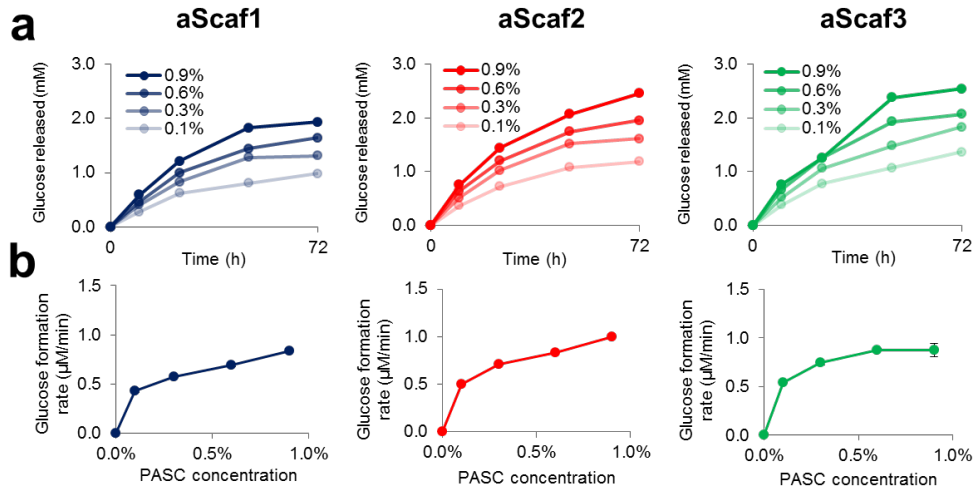


Supplementary Figure 7. Characterization of individual enzyme assembly on aScaf3-mSEA. **a** Quantification of relative individual enzyme-pScaf assembly efficiency on aScaf3-pScaf assembly. Only statistically significant difference observed was between EG and BGL (p-value 0.004). All other individual pScaf-enzyme assembly efficiencies had p-value > 0.171. **b** Binding affinity curve of the dockerin from *C. cellulovorans* that was fused to BGL and the complementary *C. cellulovorans* cohesin. A saturating amount of dockerin was incubated with yeast displaying the complementary *C. cellulovorans* cohesin at indicated concentrations until equilibrium was reached. Cells were then stained for the dockerin and analyzed using flow cytometry. Dockerin binding affinity (K_D) was determined as the concentration of dockerin at which the binding signal was half the maximum. Data are represented as the average of at least two independent experiments and error bars signify standard deviation. Statistical significance was evaluated using unpaired student t tests where ns signifies p-value > 0.05 and ** signifies p-value < 0.01.

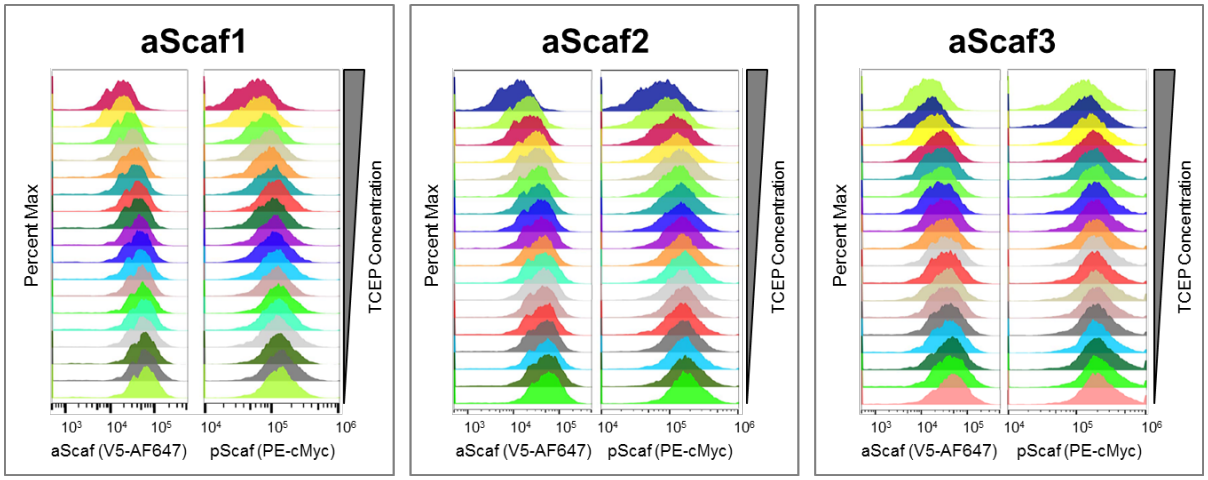


Supplementary Figure 8. Evaluating soluble enzyme and yeast surface displayed mSEA activity on PASC. **a.** Reducing-end sugar production by individual soluble enzymes and combinations of free enzymes on PASC. **b.** Degree of synergistic effect (DSE) exhibited by combinations of soluble enzymes. **c.** Reducing-end sugar production by individual enzymes and combinations of enzymes when assembled in mSEAs on aScaf1, aScaf2, and aScaf3. **d.** DSE exhibited by combinations of enzymes assembled in mSEAs on aScaf1, aScaf2, and aScaf3. Data are represented as the average of at least two independent experiments and error bars signify standard deviation. DSE was evaluated as the combined enzyme activity divided by the sum the individual enzyme activities. Activity assays were performed in 50 mM sodium acetate buffer (pH 5.0) at 30°C for 24 hours. Reducing-end sugar production was assessed using the Somogyi-

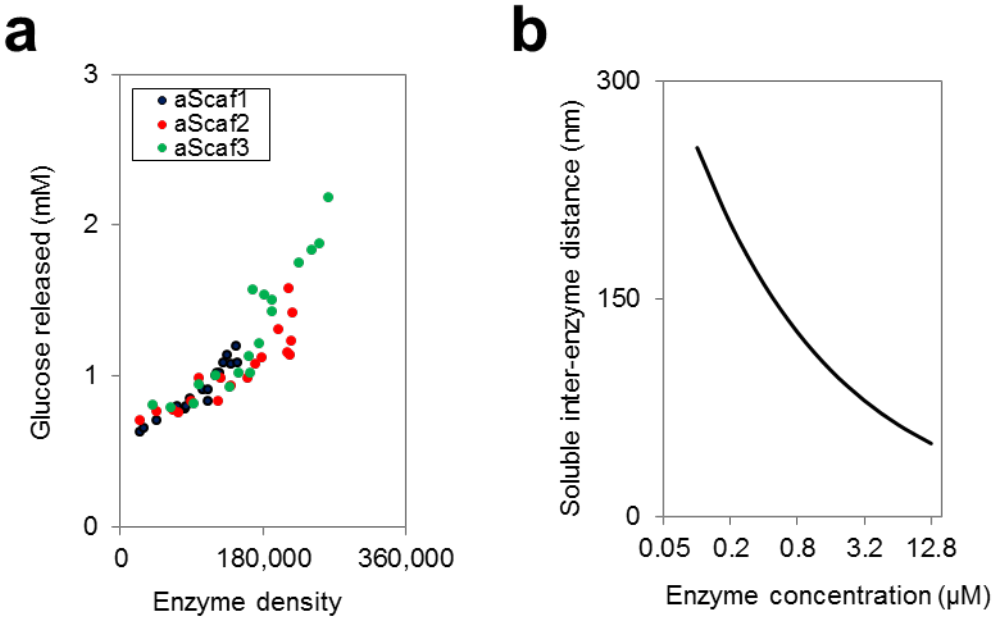
Nelson method. Data are represented as the mean of at least two independent experiments and error bars signify standard deviation.



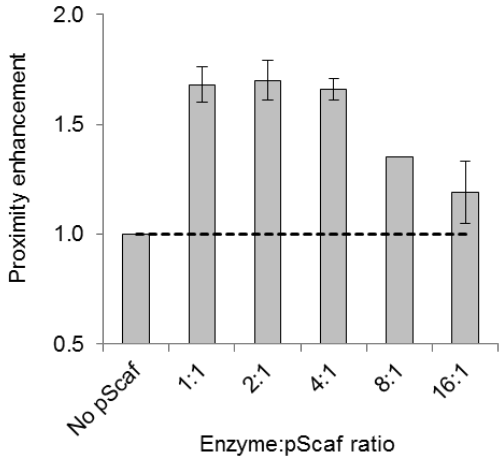
Supplementary Figure 9. Activity of yeast whole-cell biocatalysts at varying concentrations of PASC **a.** Glucose released from 0.1%, 0.3%, 0.6%, and 0.9% PASC by mSEAs assembled on aScaf1, aScaf2, and aScaf3 over 72 h. **b.** Glucose formation rate plotted with respect to PASC concentration measured after 24 h.



Supplementary Figure 10. Flow cytometry characterization of aScaf display level and pScaf assembly following varying TCEP treatments for 18-point activity assay.



Supplementary Figure 11. Evaluataing whole-cell biocatalyst performance and inter-enzyme distance for PASC hydrolysis. **a.** Glucose released from 0.3% PASC by yeast whole-cell biocatalysts displaying mSEAs on aScaf1, aScaf2, and aScaf3, plotted as a function of enzyme density. **b.** Average inter-enzyme distance for soluble enzymes plotted as a function of soluble enzyme concentration.



Supplemnnty Figure 12: Proximity enhancement on the hydrolysis of 0.3% PASC by sSEAs as a function of enzyme:pScaf ratio present in reaction mixture. Data are represented as the mean of at least two independent experiments and error bars signify standard deviation.

Supplementary Tables

Supplementary Table 1. Description of recombinant proteins used in this study.

Protein	Description	Accession number[a]	Host cell	Expression vector	Tag
Anchor scaffoldin (aScaf)	Scaffold protein containing one, two, or three Type II cohesins from scaffoldin OlpB displayed on *S. cerevisiae* surface	WP_020458018	EBY100	pYD1	c-Myc
Primary scaffoldin (pScaf)	Scaffold protein containing four of type I cohesins, one CBM, and a type II dockerin from different organisms	CipA:WP_020458017; ScaB: Q9AE52; CbpA: AAA23218; CipC: AAC28899; CBM:WP_020458017; Type II dockerin: WP_020458017	BL21 (DE3)	pET28a	V5
EG	CelA catalytic region from *C. thermocellum* with ScaA dockerin from *R. flavefaciens*	CelA catalytic region: AA83521 ScaA: CAC34384	BL21 (DE3)	pET28a	6×His
CBHI	CelF with its native dockerin from *C. cellulolyticum*	AAB41452	BL21 (DE3)	pET28a	6×His
CBHII	CelK with its native dockerin from *C. thermocellum*	AAC06139	BL21 (DE3)	pET28a	6×His
BGL	BglA with ExgS dockerin from *C. cellulovorans*	BglA: AAQ00997 ExgS: U34793	BL21 (DE3)	pET28a	6×His

[a] The accession numbers appear in the DDBJ/EMBL/GenBank nucleotide sequence databases.

Supplementary Table 2. Primers for cloning the mSEA components used in this study. Homology arms for DNA assembly[1] are shown in grey. Restriction sites are introduced where underlined.

Plasmids	Template	PCR product	Primer name	Sequence
pYD1-aScaf1, pYD1-aScaf2 pYD1-aScaf3	*C. thermocellum* genomic DNA	aScaf1, or aScaf2, or aScaf3	aScaf1-For	CGGCCGCTAGCGAAGCAACTCCAAGTATTGAAATGG
			aScaf1-Rev	TTCGAAGGGCCCGCTGGCGTCTTTTAACGGTTC
			aScaf2-Rev	TTCGAAGGGCCCATAGGAATCTGGAAGCTCTGAAGG
pYD1-Aga2-V5	pYD1	Aga2-V5	pYD1For	GTGGTTCTGCTAGCATGACTGGTGGAC
			pYD1noHisRev	GCGGGTTTAAACTCACGTAGAATCGAGACCGAGGAGAGGG
pRS415-pScaf	*C. thermocellum* genomic DNA	CipA-CBM	CipA-CBM-For	GATATCGAATTCCTGCAGCCCGGGGGATCCCTCGAGGAGCAAAAGCTCATTTCTGAAGAGGACTTGGGTGTGGTAGTAGAAATTGGC
			CipA-CBM-Rev	GCTGCACAGCGGATAAACCACCGGCACCGGGTTCTTTACCCCATACAAGAAC
	R. flavefaciens genomic DNA	ScaB	ScaB-For	GTTCTTGTATGGGGTAAAGAACCCGGTGCCGGTGGTTTATCCGCTGTGCAGC
			ScaB-Rev	CCTACTGTAACTTTAAGAGAATCGCCCTTAACAATGATAGCGCCATCAGTAAGAG
	C. cellulolyticum genomic DNA	CipC	CipC-For	CTTACTGATGGCGCTATCATTGTTAAGGGCGATTCTCTTAAAGTTACAGTAGG
			CipC-Rev	GTAGCTGTTACAGCTTCAACTGGTGTTTGAGTACCAGGATCTATAGTTACAC
	C. cellulovorans genomic DNA	CbpA	CbpA-For	GTGTAACTATAGATCCTGGTACTCAAACACCAGTTGAAGCTGTAACAGCTAC
			CbpA-Rev	TATATCCTTCTATTACAGGTTTATTGATAGTTACTGTTCCTGGGTTAACTGC
	C. thermocellum genomic DNA	Type II dockerin	DocII-For	GCAGTTAACCCAGGAACAGTAACTATCAATAAACCTGTAATAGAAGGATATAAAG
			DocII-Rev	CCACCGCGGTGGCGGCCGCTCTGAGACTGTGCGTCGTAATCACTTGATGTAG
pET28-pScaf	pRS415-pScaf	pScaf	pScaf-For	ATGGCTAGCATGGAGCAAAAGCTCATTTC
			pScaf-Rev	GTGCTCGAGTCACTGTGCGTCGTAATCAC
pET28-CBHII	*C. thermocellum* genomic DNA	CelK	CelK-For	CCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGTTGGAAGACAAGTCTTCAAAG
			CelK-Rev	AGTGCGGCCGCTTTATGTGGCAATACATC
pET28-CBHI	*C. cellulolyticum* genomic DNA	CelF	CelF- For	ATACCATGGCTTCAAGTCCTGCAAACAAG
			CelF- Rev	GGTGCTCGAGTTGGATAGAAAGAAGTGC
pET28-EG	*C. thermocellum* genomic DNA	CelA catalytic region	CelAScaA -For1	CAGCCATATGGCAGGTGTGCCTTTTAAC
			CelAScaA-Rev1	CCGGGCTGAGGAGTTGTTACACCGTAAACAACCTGAGGAG
	R. flavefaciens genomic DNA	ScaA dockerin	CelAScaA- For2	CTCCTCAGGTTGTTTACGGTGTAACAACTCCTCAGCCCGG
			CelAScaA- Rev2	GGTGCTCGAGTTACTGAGGAAGTGTGATG
pET28-BGL	*C. cellulovorans* genomic DNA	BGL	BglAExgS - For1	CCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGAAAAGCTAAGATTTCCC

			BglAExgS- Rev1	CCTGGAGTAGGGTCTGTGAACTTATTAGATCTT TCTATAAGC
		ExgS dockerin	BglAExgS- For2	GCTTATAGAAAGATCTAATAAGTTCACAGACC CTACTCCAGG
			BglAExgS- Rev2	GGTGCTCGAGAGCAAGAAGTGCTTTCTT

Supplementary Table 3: Molecular weight of each mSEA protein component.

	Size (kDa)
aScaf1	30.5
aScaf2	54.6
aScaf3	71.9
pScaf	101.2
CBHII	126.2
EG	127.5
CBHI	101.3
BGL	76.3

Supplementary References:

1. Shao, Z., Zhao, H. & Zhao, H. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res*. **37**, e16 (2009).

2. Heckman, K. L. & Pease, L. R. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat*. *Protoc*. **2**, 924–932 (2007).