

# Chimeric Streptavidins as Host Proteins for Artificial Metalloenzymes

*Michela M. Pellizzoni,<sup>1†</sup> Fabian Schwizer,<sup>1</sup> Christopher W. Wood,<sup>2</sup> Valerio Sabatino,<sup>1</sup> Yoann Cotelle,<sup>3‡</sup> Stefan Matile,<sup>3</sup> Derek N. Woolfson,<sup>2,4,5\*</sup> Thomas R. Ward<sup>1,\*</sup>*

<sup>1</sup>University of Basel, Department of Chemistry, Mattenstrasse 24a, BPR 1096, CH 4002 Basel, Switzerland.

<sup>2</sup>School of Chemistry, University of Bristol, Bristol, BS8 1TS, United Kingdom.

<sup>3</sup>School of Chemistry and Biochemistry, University of Geneva, Quai Ernest Ansermet 30, CH-1211 Geneva, Switzerland.

<sup>4</sup>School of Biochemistry, University of Bristol, Biomedical Sciences Building, University Walk, Bristol, BS8 1TD, UK

<sup>5</sup>BrisSynBio, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol, BS8 1TQ, UK

ABSTRACT The streptavidin scaffold was expanded with well-structured naturally occurring motifs. These chimeric scaffolds were tested as host for biotinylated catalysts as artificial metalloenzymes (ArM) for asymmetric transfer hydrogenation, ring closing metathesis and

anion- $\pi$  catalysis. The additional second coordination sphere elements significantly influence both the activity and the selectivity of the resulting hybrid catalysts. These findings lead to identify propitious chimeric streptavidins for future directed evolution efforts of artificial metalloenzymes.

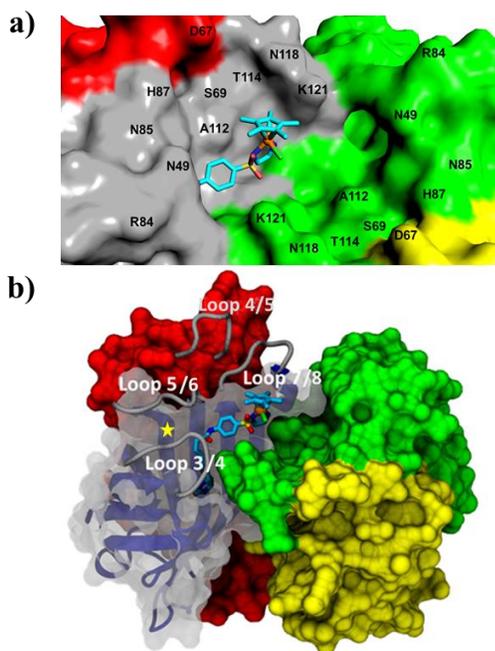
KEYWORDS Artificial metalloenzyme, chimeric protein, protein engineering, protein design, transfer hydrogenation, ring closing metathesis, anion- $\pi$  catalysis.

INTRODUCTION: Artificial metalloenzymes (ArMs hereafter), introduced by Whitesides in 1978<sup>1</sup> are aimed at combining the best features of both enzymes and homogeneous catalysts. Such ArMs result from anchoring an abiotic cofactor within a host protein.<sup>2-7</sup> This field has gained popularity thanks to the widespread use of molecular biology and recombinant protein production. Ever since our first report on ArMs based on the biotin-streptavidin technology in 2003,<sup>8</sup> we have systematically pursued a chemo-genetic optimization strategy to incrementally improve the performance of these hybrid catalysts.<sup>9-10</sup> This versatile strategy critically relies on our ability of produce streptavidin (Sav) libraries and to screen these with a variety of different biotinylated cofactors. Having identified an active ArM, twenty to fifty single-, double- and triple point Sav mutants are usually screened relying on typical directed evolution schemes. Despite the versatility of site-directed mutagenesis and directed evolution,<sup>11-13</sup> point mutations do not allow to drastically vary the topology of an active site. While such optimization efforts have proven successful (e.g. up to hundred fold increase in rate),<sup>14</sup> more diverse libraries may allow to further optimize ArMs displaying catalytic efficiencies approaching those of natural enzymes.<sup>15-</sup>

<sup>17</sup> Inspection of the ArM's structures based on the biotin-streptavidin technology highlights the

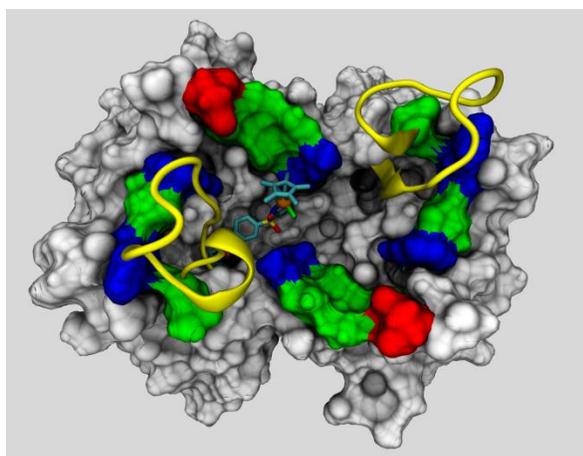
narrow dispersion of the position of metal within the biotin-binding vestibule. With few exceptions however, each ArM requires a different Sav mutant for improved performance.<sup>18-19</sup>

Inspection of the biotin-binding vestibule of Sav where the catalytic event occurs (i.e. active site) suggests that it is ideally suited to host both the cofactor and its substrate. However, Sav's lack of plasticity (as revealed by the RMSD of all C $\alpha$ 's observed in the X-ray structures of Sav and metal-loaded Sav) as well as its shallow active site, exposes the metal to the reaction medium (Figure 1a). We hypothesized that introduction of additional structural features may offer the opportunity to further tailor and confine the active site. With this goal in mind, we set out to engineer Sav by introducing additional structural motifs around the biotin-binding vestibule (Figure 1b). The chimeras presented herein were designed to enable the exploration of diverse topologies around the active site by the introduction of additional second coordination-sphere elements.<sup>20</sup>



**Figure 1.** Close-up view of the X-ray structure of homotetrameric Sav S112A with one biotinylated cofactor [Cp\*Ir(biot-*p*-L)Cl] bound (PDB ID 3PK2).<sup>21</sup> The surface representation displays a single cofactor (stick representation; Ir, orange sphere, Cl, green) and close-lying amino acid residues. The shallowness of the biotin-binding vestibule (i.e. the active site) suggests that engineering additional structural motifs may provide additional second coordination sphere elements around the biotinylated cofactor a). Surface display representation of the Sav tetramer. The Sav loops that were engineered to create chimera proteins are labeled: loop 3/4, residues 46 – 52; loop 4/5, residues 63 – 70; loop 5/6, residues 82 - 87 and loop 7/8, residues 113 - 117). The yellow star highlights the position selected for the creation of Chimera MP\_Sav family, b). Monomer surfaces are highlighted in green, yellow, red and (translucent) grey, respectively.

Thus, we set out to modify the ArMs genetically to introduce well-defined secondary structures around the “active site”. Thanks to the dimer-of-dimer nature of homotetrameric Sav, genetic modifications are reflected twice in each of the two active sites (Figure 2) and thus can be anticipated to have a significant shielding effect on the Sav-embedded cofactor.



**Figure 2.** Cartoon representation of a homotetrameric Sav (surface representation) construct bearing an additional motif (entry 8 in Table 1) engineered in loop 3,4 (yellow cartoon representation). The model for the chimeric protein was generated from PDB ID 3PK2, using

homology modeling and structure refinement with Yasara. (color code: red = acidic, blue = basic, green = polar, white = apolar)

## RESULTS AND DISCUSSION

**Protein design.** Our initial efforts focused on inserting (GGX)<sub>n</sub> repeat sequences (where X = S, N, G, V, I, H, R and n = 1-8) at various loop positions that make up the biotin-binding vestibule. We selected the following positions for the insertion: G48–N49, T66–D67, R84–N85 and A117–N118 to introduce one or two (GGX)<sub>n</sub> repeat motifs ( Table 1).

**Table 1** Summary of the expressed Sav constructs bearing (GGX)<sub>n</sub> repeat motifs.

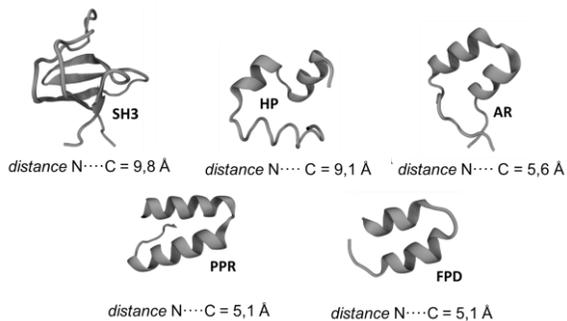
entry	Sav loop	loop sequence	number of (GGX) <sub>n</sub> repeats	Expression level <sup>a</sup> (mg/l)
1	-	G48–N49	0	120
2	3/4	G48–GGSGGS–N49	2	13.8
3	4/5	T66–GGSGGS–D67	2	16.3
4	5/6	R84–GGSGGS–N85	2	13.0
5	7/8	A117–GGSGGS–N118	2	not purified
6	3/4 + 4/5	G48–GGSGGS–N49 T66–GGSGGS–D67	4	1.5
7	3/4 + 5/6	G48–GGSGGS–N49 R84–GGSGGS–N85	4	0.9
8	3/4 + 7/8	G48–GGSGGS–N49 A117–GGSGGS–N118	4	not purified
9	3/4	G48–GGNNGGNGGGGGVGGGS–N49	5	55.0
10	3/4	G48–GGIGGSGGGGGHGGRGGGGGVGGGS–N49	8	not purified
11	3/4	G48–GGNNGSGGGGGGGSGGS–N49	7	not purified
12	3/4	G48–GGRGGGGGHGGCGGVGGGS–N49	6	not purified

<sup>a</sup> Expressed as soluble and functional (i.e. binds biotin-4-fluorescein) tetrameric fraction. The expression conditions are detailed in the SI.

Having identified Sav positions tolerant to (GGX)<sub>n</sub> insertions, we next sought to replace these GGX repeats, which we presumed would be unstructured, by naturally occurring motifs with a

well-defined secondary structure. We pursued the construction of two different chimera streptavidin families: a) streptavidin containing an extended 2D structural motif (24-60 amino acid residues (aa hereafter) with  $\alpha$  helix or  $\beta$  sheets secondary structure, 2D\_Sav, hereafter, Table 2) and b) streptavidin containing shorter naturally-occurring loops (5-12 amino acid residues, MP\_Sav, hereafter, Table 3).

The first family with additional structured elements around the biotin-binding vestibule, 2D\_Sav, was created by fusion of Sav with highly conserved- and structured peptide sequences (see SI) that display close-lying *N*- and *C*-termini. For this purpose, the following sequences were selected: SH3 (Src Homology 3 domain, 60 aa) containing antiparallel  $\beta$ -sheets,<sup>22</sup> AR (Ankyrin, 30-34 aa),<sup>23</sup> HP (Villin Headpiece subdomain, 35 aa),<sup>24</sup> FPD (FoldIt Players Design, 24 aa)<sup>25</sup> and PPR (Penta-trico-Peptide Repeat 35 aa)<sup>26</sup> containing an helix-turn-helix motif (Figure 3 and Table 2). Inspection of the X-ray structure of *mature* “apo”-Sav, which bears an *N*<sub>term</sub>-T7-solubility tag at positions 1 - 12 and extends to position 159,<sup>27</sup> revealed that the *C*<sub>term</sub> occupies the biotin-binding site.<sup>28</sup> We thus introduced at the *C*<sub>term</sub> either the SH3 or PPR consensus sequence to provide an additional well-structured secondary structure, potentially close to biotin-binding vestibule. The SH3, AR, HP and FDP motifs were introduced in four different loop regions of Sav, highlighted in Figure 1b, taking into consideration their *N*<sub>term</sub> – *C*<sub>term</sub> distance. Thus, nineteen chimeric Sav genes were constructed (Figure 3 and Table 2).



**Figure 3.** Cartoon-representation of highly conserved- and structured peptide motifs introduced in Sav loops (see also Figure 1b) to generate 2D\_Sav family. The inserted structures with  $N_{term}$ - $C_{term}$  distances (N·····C) are displayed as grey cartoon representation. SH3 (Src Homology 3 domain, ca. 60 aa, antiparallel  $\beta$ -sheets);<sup>22</sup> AR (Ankyrin, 30-34 aa, helix-turn-helix);<sup>23</sup> HP (Villin Headpiece subdomain, 35 aa, helix-turn-helix);<sup>24</sup> FPD (FoldIt Players Design, 24 aa, helix-turn-helix)<sup>25</sup> and PPR (Penta-trico-Peptide Repeat 35 aa, helical hairpin).<sup>26</sup>

**Table 2.** Chimeric Proteins (2D\_Sav's family) expressed recombinantly in this study. Each protein is abbreviated with the acronym of the introduced motif; the number indicates the position of the insertion in the original Sav sequence.<sup>a</sup>

<b>entry</b>	<b>abbreviation</b>	<b>solubility</b>
1	PPR_159	Yes
2	SH3_46-52	After refolding
3	SH3_63-70	No
4	SH3_82-87	No
5	SH3_113-117	After refolding
6	SH3_159	Yes
7	AR_46-52	After refolding
8	AR_64-70	After refolding
9	AR_81-84	No
10	AR_115-117	After refolding
11	HP_46-52	Yes
12	HP_63-70	After refolding
13	HP_64-70	Yes
14	HP_82-84	Yes
15	HP_113-117	No
16	FDP_46_52	Yes
17	FDP_64-70	After refolding
18	FDP_81-84	Yes
19	FDP_115-117	After refolding

<sup>a</sup>Expression and refolding conditions are detailed in the SI.

The second family of chimera, MP\_Sav, was generated by selecting natural loops with appropriate  $N_{term}$ - $C_{term}$  distances for the substitution of Sav's residues between A46 and A50 (loop 3/4, Figure 1b). Three generations of chimeric Savs were derived from the MP variants. These were produced, purified and tested in catalysis (Table 3). The first generation included the exact sequence extracted using the ISAMBARD algorithm (Table 3, third column). Based on extensive experience with ArMs based on the biotin-streptavidin technology,<sup>19</sup> we hypothesized that the cationic residues introduced in the immediate proximity of the cofactor may be detrimental to catalytic performance. Accordingly, we mutated the Lys and Arg residues in the

second generation of MP variants (Table 3, 4-5<sup>th</sup> columns). Finally, a third generation was generated for the MP3 construct that included the highly beneficial K121F mutation, combined with the lysine residue in the engineered loop.

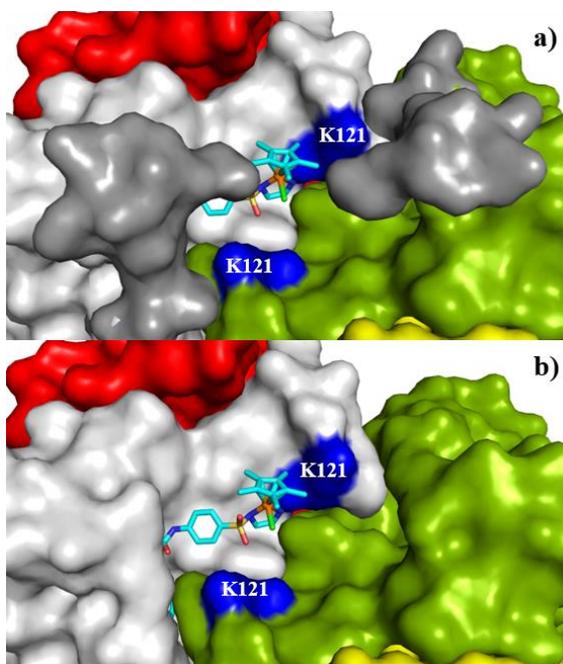
To generate initial models of the chimeric proteins, a database of loop regions from X-ray protein crystal structures from the Protein Data Bank<sup>29</sup> was created using ISAMBARD.<sup>30</sup> Loops were defined as any continuous region of backbone containing any mixture of random coil, hydrogen-bonded turn isolated beta-bridge or bend, as identified by DSSP.<sup>31-32</sup> Tools within the AMPAL module of ISAMBARD were then used to extract the backbone structure of the loop, which was stored in a database along with the PDB ID, chain and residue labels, surrounding regions of secondary structure, sequence, end-to-end distance (i.e. the distance between the C<sub>α</sub> of the residue preceding the loop and the C<sub>α</sub> of the residue immediately after the loop) and resolution of the X-ray structure. Redundant sequences were *not* removed from the database to allow any conformational diversity of the loops to be captured, with the rationale that different structures of the same, or homologous proteins, might contain very different loop conformations due to their highly flexible nature. Candidate loop designs were identified to span between residue A46 and A50 of Sav (loop 3/4), based on an X-ray protein crystal structure (PDB ID 3PK2).<sup>21</sup> Initially, these were initially filtered purely based on the end-to-end distance of the loop, requiring it to be within 0.5 Å of the A46 - A50 distance in 3PK2. Loops satisfying this criterion were fitted by aligning the backbone atoms of the residues immediately before and after the loop with those of residues A46 and A50 respectively. The root-mean square deviation (RDMS) of the distances between complementary backbone atom pairs was calculated to evaluate the quality of the fit (detailed in the SI). Models were sorted based on this quality and the candidate loops were evaluated manually considering the diversity of number and nature of amino acids, symmetry of the host and proximity to the metal center. On this basis, twelve loops

were selected for insertion into Sav replacing the wild type sequence between A46 and A50 (Table 3 and Figure 4). Subsequently, the family members of MP\_Sav's were subjected to a first round of mutagenesis, in which cationic arginine and/or lysine residues in the loop sequences, which are known to be detrimental to catalysis,<sup>33</sup> were mutated to either alanine or phenylalanine. Finally, a K121F mutation was introduced in the first and second generation of MP 3 constructs in view of its improved catalytic performance with various ArMs based on the biotin-streptavidin technology.<sup>34</sup>

**Table 3.** Selected chimera of MP\_Sav family containing natural loops with commensurate  $N_{term}$ – $C_{term}$  topologies inserted between A46 and A50 of Sav. ISAMBARD was used to create a database of loops regions from known X-ray crystal structures.<sup>a</sup>

entry	abbreviation	inserted sequence <sup>a</sup>	abbreviation	second generation MP_Sav <sup>a</sup>
1	WT	AVGNA <sup>b</sup>		
2	MP 1	GKTKG	MP 1 <i>K-F</i> <sup>d</sup>	GATFG
3	MP 2	GRSRG		–
4	MP 3	GNLKYG <sup>c</sup>	MP 3 <i>K-A</i> <sup>d</sup>	GNLAYG <sup>c</sup>
			MP 3 <i>K-F</i> <sup>d</sup>	GNLFYG
5	MP 4	GIDRNG		–
6	MP 5	GDMKPRG		–
7	MP 6	GHEKRDG	<i>MP 6 K-A_R-F</i> <sup>d</sup>	GHEAFDG
8	MP 7	GKHNKPDDCG		–
9	MP 8	GRRQIGTRSG		–
10	MP 9	GEPFGGEKING	MP 9 <i>K-F</i> <sup>d</sup>	GEPFGGEFING
11	MP 10	GGRVIPVKLGG		–
12	MP 11	GYLSSQNGQPG		–
13	MP 12	GTERPSKNSHPG	MP 12 <i>R-A_K-F</i> <sup>d</sup>	GTEAPSFNSHPG

<sup>a</sup> The expression and refolding conditions are detailed in the SI. <sup>b</sup> This sequence was deleted from the WT in all MP constructs. <sup>c</sup> The third generation of MP constructs was generated combining this loop sequence and the K121F mutation. <sup>d</sup> Corresponds to the point mutation(s) introduced in the corresponding first generation MP construct to remove cationic residues.



**Figure 4.** Model of MP\_Sav member (MP\_11 depicted), containing the loop GYLSSQNGQPG insert (dark grey) between positions A46 and A50 generated from PDB ID 3PK2. The insert partially shields the biotin-binding vestibule, a). For comparison, the surface representation of PDB ID 3PK2 is displayed in the same orientation, b). Monomer surfaces are highlighted in green, yellow, red and light grey, respectively. The residues K121 are highlighted in blue.

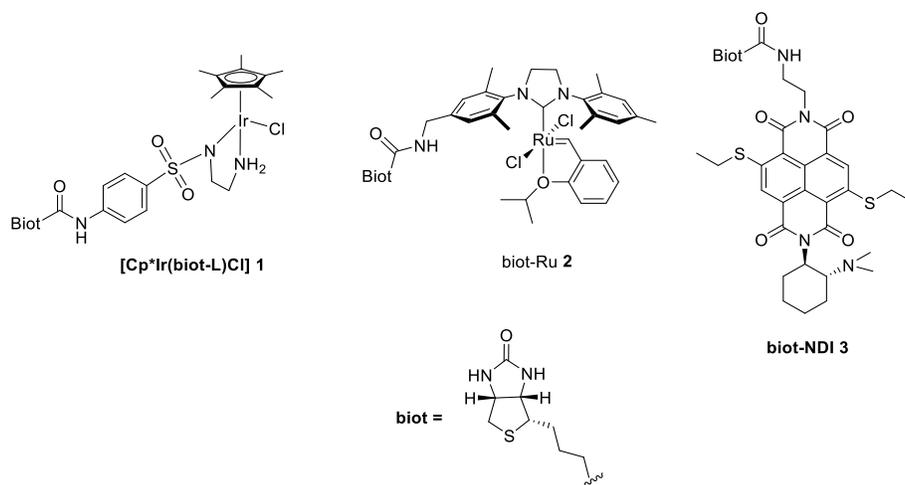
**Protein overexpression.** *E. coli* BL21(DE3) containing the pET24 (+) plasmid was used for the overexpression of all chimera streptavidins using an autoinduction medium.<sup>35</sup> After cell lysis, SDS gel analysis of the supernatant was used to confirm the overexpression and the biotin-binding capacity of the fusion proteins. For this purpose, biotin-4-fluorescein (B4F) was added to the Sav chimera. The presence of fluorescent bands in the corresponding SDS PAGE confirmed the proper folding and functionality of the Sav constructs. While the majority of constructs containing the shorter loops resulted in soluble and properly-folded proteins (75%, see Table S3), the 2D\_Sav family members displayed different folding behaviors (only 40% were soluble after overexpression, see Table 2 and Table S3) depending on: i) the type of motif used and ii)

the site of insertion. Proteins containing PPR and SH3 at the *C*-terminus were usually well folded as well as Sav containing either HP and FDP motifs engineered in either loop 3/4 or 5/6. As summarized in Table 2, some Sav constructs were produced as inclusion bodies however.

Screening various refolding buffers revealed that the majority of insoluble chimeras could be refolded using a highly dilute MES buffer (see SI, Table S3 entries 1-19). When required, this step allowed to isolate soluble and biotin-binding chimeric Sav. Eight of twelve insoluble 2D\_Sav constructs (Table 2) could be purified using a biotin-sepharose affinity column (yields and purification methods detailed in Table S3). After purification, all the constructs were characterized by SDS PAGE and mass-spectroscopy (See SI).

The introduction of extended loops around the active site of the ArMs does not significantly affect the stability or the aggregation state of the engineered proteins. Indeed, none of chimeric proteins are denatured under the "denaturing" SDS PAGE conditions: the chimeric proteins migrate as tetramers and maintain their biotin-4-fluorescein binding capacity.

**Catalytic performance.** Next, chimeric Sav were combined with various biotinylated cofactors and tested as artificial metalloenzymes in the following catalytic reactions: Asymmetric Transfer Hydrogenation (ATH), Ring Closing Metathesis (RCM), and anion  $\pi$ -catalysis (ANPI). The corresponding biotinylated catalysts are displayed in Figure 5.



**Figure 5.** Biotinylated cofactors used to complete the Sav for: i) asymmetric transfer hydrogenation reaction using  $[\text{Cp}^*\text{Ir}(\text{biot-L})\text{Cl}]$  **1**, ii) ring closing metathesis using a biotinylated Hoveyda–Grubbs second-generation catalyst (biot-Ru **2** hereafter) and iii) anion- $\pi$  catalysis on a biotinylated naphthalenediimide surface (**biot-NDI 3** hereafter).

Previous work revealed that the Ir- $d^6$  piano-stool  $[\text{Cp}^*\text{Ir}(\text{biot-L})\text{Cl}]$  **1** was the most active ATH catalysts for the reduction of prochiral imines in the presence of Sav.<sup>36-37</sup> Two different prochiral cyclic imines **4** and **6** were selected and the results of the asymmetric transfer hydrogenation using  $[\text{Cp}^*\text{Ir}(\text{biot-L})\text{Cl}]$  **1**, are collected in Table 4. Compared to the free cofactor  $[\text{Cp}^*\text{Ir}(\text{biot-L})\text{Cl}]$  **1** (Table 4 entries 1 and 2) and the corresponding benchmark ATHase  $[\text{Cp}^*\text{Ir}(\text{biot-L})\text{Cl}] \cdot \text{WT Sav}$  (Table 4 entries 3 and 4), some of the engineered Sav constructs performed better, both in terms of TON (turnover number) and enantioselectivity. While the improvement of enantioselectivity was modest at best, the TON number could be improved by up to a factor two for substrate **4** (Table 4 entries 5 and 6) and up to a factor seven for substrate **6** (Table 4 entries 7 and 8). Importantly, the position of the insertions plays a critical role in the activity of the corresponding ATHase. Insertion in loop 3/4 (i.e. positions 46 - 52) affords the most active ATHases, for both HP\_46-52 and FDP\_46-52 constructs, Table 4. For comparison, introduction of the HP in the loop 4/5 (i.e. positions 64 - 70) had a negative impact on the ATHase activity

(Table 4 entry 9). We hypothesize that the insertion between positions 46 - 52 projects the loop in the proximity of the Ir-center and thus has the highest (positive) impact on the ATHase's performance. While the improvement in selectivity remains modest, the activity is significantly affected, resulting in up to a sevenfold increase in turnovers after sixteen hours.

**Table 4.** Selected ATHase results obtained with the 2D\_Sav combined with [Cp\*Ir(biot-L)Cl]

1.<sup>a</sup> Error margins in square brackets [+/-].

entry	2D_Sav	substrate	ee ( <i>R</i> )	TON	catalytic improvement vs WT Sav
1	No protein	<b>4</b>	0 [nd]	197 [0.1]	-
2	No protein	<b>6</b>	0 [nd]	0 [nd]	-
3	WT Sav	<b>4</b>	45 [0.5]	170 [2.0]	-
4	WT Sav	<b>6</b>	76 [1.2]	22 [0.7]	-
5	HP_46-52	<b>4</b>	59 [0.6]	223 [2.2]	1.7 <sup>b</sup>
6	FDP_46-52	<b>4</b>	51 [1.0]	340 [5.1]	2.6 <sup>b</sup>
7	HP_46-52	<b>6</b>	80 [1.8]	162 [2.3]	7.3 <sup>c</sup>
8	FDP_46-52	<b>6</b>	82 [2.7]	158 [2.5]	7.2 <sup>c</sup>
9	HP_64-70	<b>4</b>	22 [2.0]	14 [1.2]	0.01 <sup>c</sup>

<sup>a</sup> reaction conditions: substrate 10 mM, 16 h, 37 °C, [Cp\*Ir(biot-L)Cl] **1** 10 μM, Sav's FBS 20 μM, MOPS 0,6 M, pH = 7,5, formate 3 M, V<sub>tot</sub> = 200 μL. <sup>b</sup> Catalytic improvement based on initial rates (after 1 h reaction) calculated by the ratio between TON 2D\_Sav and TON of WT Sav. <sup>c</sup> Catalytic improvement calculated from the ratio between TON 2D\_Sav and TON of WT Sav (after 16 h).

Considering the MP\_Sav constructs, a general trend is observed for the reduction of substrate **4**: While the first-round Lys/Arg-containing loop sequences had dramatically lower activities but comparable selectivities to the benchmark ATHase [Cp\*Ir(biot-L)Cl] **1** · WT Sav (Table 5 entries 3 vs entries 6 – 10), substitution of of Lys to Ala or Phe within the engineered loops resulted in second-generation hybrid catalysts that rivaled or outperformed the WT ATHase (Table 5, entries 11 – 12). The [Cp\*Ir(biot-L)Cl] **1** · Sav K121F ATHase displayed significantly improved TON although this was at the cost of ee (13% ee (*S*)-**5**, 970 TON and 45% ee (*R*)-**5**, 170 TON for Sav K121F and WT, respectively). Combining the beneficial K121F mutation and loops, with or without a Lys residue, in all cases affords ATHase with lower TONs compared to Sav K121F (Table 5, entries 13 – 14). For substrate **6**, all second-generation MP\_Sav performed better than the corresponding ATHase [Cp\*Ir(biot-L)Cl] **1** · WT Sav (Table 5 entry 4 vs Table 5 entries 15 – 17). However [Cp\*Ir(biot-L)Cl] **1** · Sav K121F followed by the chimera [Cp\*Ir(biot-L)Cl] **1** · Sav HP\_46-52 and [Cp\*Ir(biot-L)Cl] **1** · Sav FDP\_46-52 were the most active ATHases (Table 5 entry 5 and Table 4 entries 7 and 8). A summary of all ATHase experiments is presented in SI Tables S4 and S5.

**Table 5.** Selected ATHase results obtained for the MP\_Sav introduced between position A46 and A50 combined with [Cp\*Ir(biot-L)Cl] **1**.<sup>a</sup> Error margins in square brackets [+/-].

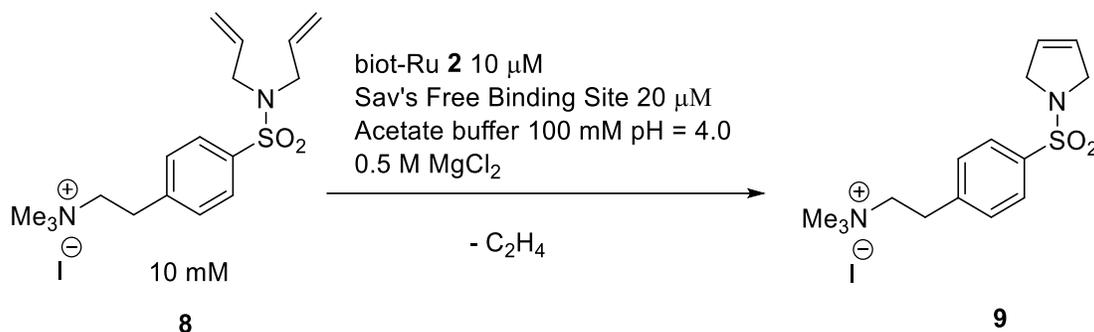
entry	MP_Sav <sup>b,c</sup>	substrate	ee ( <i>R</i> )	TON	catalytic improvement vs WT Sav
1	No protein	<b>4</b>	0	197[0.2]	-
2	No protein	<b>6</b>	0	0	-
3	WT Sav	<b>4</b>	45 [0.5]	170 [2.0]	-
4	WT Sav	<b>6</b>	76 [1.2]	22 [0.7]	-
5	Sav K121F	<b>4</b>	-13 ( <i>S</i> ) [0.1]	970 [8.7]	10 <sup>d</sup>
6	MP 3	<b>4</b>	53 [2.3]	6 [0.2]	0.04 <sup>e</sup>
7	MP 4	<b>4</b>	42 [0.1]	67 [2.5]	0.4 <sup>e</sup>
8	MP 9	<b>4</b>	53 [0.5]	82 [3.4]	0.48 <sup>e</sup>
9	MP 11	<b>4</b>	45 [0.4]	61 [5.5]	0.36 <sup>e</sup>
10	MP 12	<b>4</b>	62 [0.2]	15 [1.6]	0.08 <sup>e</sup>
11	MP 3_ <i>K-A</i>	<b>4</b>	46 [0.9]	192 [3.0]	1.75 <sup>d</sup>
12	MP 9_ <i>K-F</i>	<b>4</b>	49 [0.4]	198 [2.1]	1.65 <sup>d</sup>
13	MP 3_K121F	<b>4</b>	15 [0.6]	568 [9.3]	3.7 <sup>d</sup>
14	MP 3_ <i>K-A</i> -K121F	<b>4</b>	14 [0.1]	462 [4.3]	3 <sup>d</sup>
15	MP 3_ <i>K-A</i>	<b>6</b>	79 [1.6]	57 [2.5]	2.5 <sup>e</sup>
16	MP 6_ <i>K-A</i> _R-F	<b>6</b>	77 [1.5]	100 [1.9]	4.5 <sup>e</sup>
17	MP 9_ <i>K-F</i>	<b>6</b>	80 [2.0]	71 [1.0]	3.2 <sup>e</sup>

<sup>a</sup> reaction conditions: substrate 10 mM, 16 h, 37°C, [Cp\*Ir(biot-L)Cl] **1** 10 μM, Sav's Free Binding Site (FBS) 20 μM, MOPS 0,6 M, pH = 7,5, formate 3 M, V<sub>tot</sub> = 200 μL. <sup>b</sup> The numbering of the residues of Sav was kept identical to the WT numbering, even though the insertion may be placed before the position of the mutation. <sup>c</sup> Italicized one letter aminoacid abbreviations designate cationic aminoacids within the inserted loops that were mutated to either A or F to probe the effect of the charge on the catalytic performance (See Table 3 for details). <sup>d</sup> Catalytic improvement based on initial rates (after 1 h reaction) calculated by the ratio between TON MP\_Sav and TON of WT Sav. <sup>e</sup> Catalytic improvement calculated by the ratio between TON 2D\_Sav and TON of WT Sav (after 16 h).

Qualitative kinetic data for substrate **4**, reveal that the turnover frequency (TOF) is significantly affected by the introduction of loops. After one hour, the TOF for FDP\_46-52 approaches 50 hour<sup>-1</sup>, whereas the TOF barely reaches 20 hour<sup>-1</sup> for WT. Similar trends are observed after 16 hours, see SI Fig. S4. We speculate that this effect reflects the interaction of the loops with the catalyst, rather than merely stabilizing the Sav host. Unfortunately, all

attempts to crystallize the chimeric ArMs revealed that the expanded loops were disordered, the rest of the protein being well-ordered.

Building on our previous work on olefin metathesis in aqueous phase,<sup>38</sup> we investigated the catalytic behavior of 2D\_Sav and MP\_Sav scaffolds in the presence of biot-Ru **2**. The resulting artificial metathases were tested for the RCM of substrate **8**. Both the biot-Ru **2** · Sav HP 64-70 and biot-Ru **2** · Sav PPR 159 outperform the bare catalyst (Table 6, entries 1 vs 4 and 5). As for the ATHases, introduction of a Phe residue at position K121 improved catalytic performance resulting in 105 TONs (Table 6, entry 3). The second and the third generation MP\_Sav variants originating from the Sav MP3 sequence displayed improved activity. (Table 6 entries 6-8). However, compared to biot-Ru **2** · Sav K121F, none of the newly designed chimeric Sav outperformed this single point mutant (Table 6 entries 4-8). A complete list of chimeric metathases is collected in SI Table S6.

**Table 6.** Selected results for RCM of diallyl substrate **8** using 2D\_Sav and MP\_Sav.<sup>a</sup>

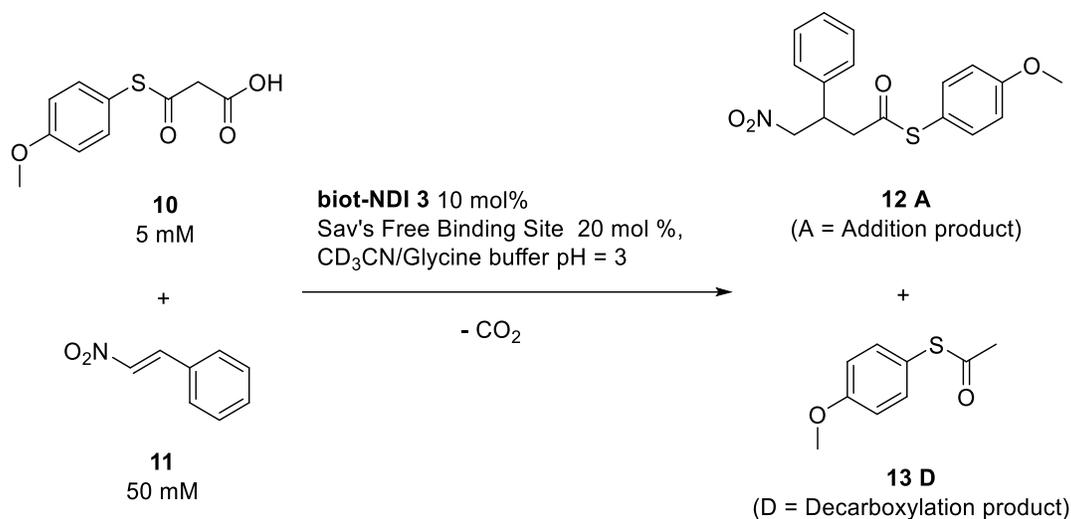
entry	Sav <sup>b,c</sup>	TON
1	No protein	38 [1.5]
2	WT Sav	87 [1.1]
3	K121F	105 [0.1]
4	HP_64-70	56 [0.5]
5	PPR_159	64 [0.7]
6	MP 3_K-A	61 [0.1]
7	MP 3_K121F	88 [1.6]
8	MP 3_K-A-K121F	98 [3.0]

<sup>a</sup> reaction conditions: substrate **8** 10 mM, biot-Ru **2** 10 μM, 16 h, 37°C, Chim\_Sav and MP\_Sav 20 μM Free Binding Site, acetate buffer 100 mM, 0.5 M MgCl<sub>2</sub> pH = 4.0, V<sub>tot</sub> = 200 μL. <sup>b</sup> The numbering of the residues of Sav were kept as the WT numbering, even though the insertion may be placed before the position of the mutation. <sup>c</sup> Italicized one letter aminoacid abbreviations designate cationic aminoacids within the inserted loops that were mutated to either A or F to probe the effect of the charge on the catalytic performance (See Table 3 for details).

The Michael addition of malonic acid half thioester **10** (MAHT) was investigated next. Importantly, anion- $\pi$  interactions were shown to have a significant impact on the chemoselectivity of this reaction by selectively producing the disfavored addition product **12 A**. Without an electron-deficient  $\pi$ -system as the catalyst, the decarboxylation product **13 D** is generated almost exclusively.<sup>39-40</sup> By linking a biotin to an NDI  $\pi$ -surface and incorporation within Sav anion- $\pi$  enzymes result. These artificial enzymes produce the addition product **12 A** with an exquisite chemoselectivity (A/D ratio > 30) and enantioselectivity (ee = 95% for biot-NDI **3**· Sav S112Y)<sup>41</sup> through synergism between NDIs and Sav mutants. However, due the inhibition by anions, the use of a more confined cavity may be of interest. With this goal in

mind, the catalysis was tested with the previously established conditions.<sup>41</sup> Most of the chimeric Sav's produced the addition product **12 A** with excellent chemoselectivity (**12 A**/**13 D** > 30) and moderate yields (up to 61 Yield%, Table 7 entry 2). Three of the chimeric Sav's were inactive. We posit that the lack of reaction may be attributed to an important shielding above the NDI surface embedded in either MP 12 Sav, Sav FDP\_46-52 and Sav HP\_46-52. Only **biot-NDI 3** · Sav PPR\_159, containing the helical hairpin domain at the *C<sub>term</sub>* afforded **12 A** with an enantioselectivity (Table 7, entry 3), thus competing with the most efficient anion- $\pi$  organocatalysts achieved to date.<sup>40, 42</sup> A summary of all anion- $\pi$  catalytic experiments is presented in SI, Table S7.

**Table 7.** Selected results obtained for anion- $\pi$  catalysis using **biot-NDI 3** · 2D\_Sav or **biot-NDI 3** · MP\_Sav.<sup>a</sup>



entry	Sav <sup>b,c</sup>	yield %	12 A/13 D	ee (%)
1	WT Sav	60	>30	41
2	MP 12_ <i>R-A_K-F</i>	61	>30	0
3	PPR_159	57	>30	22
4	MP 3_ <i>K-A_K121F</i>	55	>30	0
5	MP 3_K121F	51	>30	0
6	MP 9	23	>30	0
7	MP 9_ <i>K-F</i>	45	>30	0
8	MP 12	traces	nd	0
9	FDP_46-52	0	nd	nd
10	HP_46-52	0	nd	nd

<sup>a</sup> reaction conditions: **biot NDI 3** 10 mol%, Sav's FBS 20 mol %, CD<sub>3</sub>CN/Glycine buffer pH = 3, substrate **10** 5 mM, substrate **11** 50 mM. <sup>b</sup> The numbering of the residues of Sav were kept as that of WT Sav, even though the insertion may be placed before the position of the mutation. <sup>c</sup> Italicized one letter aminoacid abbreviations designate cationic aminoacids within the inserted loops that were mutated to either A or F to probe the effect of the charge on the catalytic performance (See Table 3 for details).

## OUTLOOK

In summary, this work highlights that the loop 3/4 and the *C*-terminus are the most promising regions to engineer the next generation of Sav-based ArMs using directed evolution protocols. The introduction of additional motifs generates soluble chimeric Sav without significantly affecting their biotin binding capability. The additional structured elements around the biotin-binding vestibule were created by fusion of Sav with highly conserved- and structured peptide sequences displaying close-lying *N*- and *C*-termini or applying rational loop mutagenesis techniques using ISAMBARD. In particular, this study revealed that FDP\_46-52, HP 46-52, PPR\_159 and MP3 members are promising candidates thanks to their good expression yields of and their impact on ATH, RCM and anion  $\pi$  catalysis.

The catalytic performance of ArMs derived from the chimeric Savs described herein suggest that these straightforward modifications of the Sav host has a significant impact on the catalytic activity. Thanks to the remarkable stability of the Sav scaffold, introduction of additional structural motifs leads, in the vast majority of cases, to soluble and functional chimeric Sav. For the ATHase, the resulting activity was up to sevenfold higher than the parent WT Sav. Thus, this strategy of embellishing loops proximal to the active site offers a versatile means to complement directed evolution efforts to optimize the performance of the ArMs based on the biotin-streptavidin technology. The newly introduced structural motifs might be further engineered to form a protective lid over the active site, resulting in ArMs able to perform catalytic reactions in cell free extracts or in the cytoplasm.

## EXPERIMENTAL SECTION

**General procedure for the asymmetric transfer hydrogenation.** 10  $\mu$ l of proteins stock solution in Milli-Q H<sub>2</sub>O (200  $\mu$ M free binding site), was added to 185  $\mu$ l of reaction buffer followed by the addition of 5  $\mu$ l of the biotinylated metal complex [Cp\*Ir(biot-L)Cl] **1** from a

stock solution (0.4 mM in DMSO). The solution was mixed for 20 min at 37 °C and 800 rpm in a thermo-mixer for precomplexation. Finally, 5 µl of substrate **4** or **6** stock solution (400 mM in DMSO) was added and the mixture was stirred at 37 °C for 16 h. Subsequently, 20 % NaOH solution was added to the reaction mixture, followed by the addition of CH<sub>2</sub>Cl<sub>2</sub>. After mixing, the organic phase was separated and dried with anhydrous Na<sub>2</sub>SO<sub>4</sub>. Solids were separated through centrifugation and the supernatant was analyzed by using HPLC or GC.

**General procedure for ring closing metathesis.** In a 1.2 mL glass vial, 10 µl of protein stock solution (400 µM) and 5 µl biot Ru **2** (400 µM) were added to 85 µl of reaction buffer and incubated at 37°C for 20 min. After incubation, 100 µl of substrate 20 mM were added to the reaction mixture and the reaction was stirred at 37°C for 16 hours at 1000 rpm. After reaction, methanol and the internal standard were added to the reaction mixture and the whole volume transferred to Eppendorf tubes for centrifugation. The supernatant was then transferred into HPLC vials containing MQ water and the sample analysed by UPLC-MS for the quantification of product **9**.

**General procedure for anion- $\pi$  catalysis.** Stock solutions of substrates **10** (40 mM), **11** (400 mM) and biot NDI **3** (2 mM) were prepared in CD<sub>3</sub>CN. Solutions of substrates **10** should be freshly prepared as decarboxylation to afford **13 D** builds-up on aging. Solutions were prepared by mixing successively Sav WT or mutants, biotinylated ligand **3**, substrates **10**, **11** and stirred at 20 °C. After 24 h, the mixture was extracted with CDCl<sub>3</sub>, dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and analyzed by <sup>1</sup>H-NMR spectroscopy.

AUTHOR INFORMATION

**Corresponding Authors**

\*E-mail: thomas.ward@unibas.ch

D.N.Woolfson@bristol.ac.uk

### **Present Address**

† Adolphe Merkle Institute, University of Fribourg, Chemin des Verdiers 4, 1700 Fribourg, Switzerland.

‡ University of Basel, Department of Chemistry, Mattenstrasse 24a, BPR 1096, CH 4002 Basel, Switzerland.

### **Notes**

The authors declare no competing financial interest.

ACKNOWLEDGMENT Generous funding for this research was provided by the Swiss National Science Foundation (grant 200020 162348 to TRW) as well as the NCCR Molecular Systems Engineering (to SM and TRW). Precious metal salts were kindly provided by Umicore.

### ASSOCIATED CONTENT

**Supporting Information.** The following files are available free of charge.

Supplementary natural loop design, cloning procedure, protein expression and purification, analytical methods (PDF)

## REFERENCES

1. Wilson, M. E.; Whitesides, G. M., *J. Am. Chem. Soc.* **1978**, *100*, 306-307.
2. Heinisch, T.; Pellizzoni, M.; Durrenberger, M.; Tinberg, C. E.; Kohler, V.; Klehr, J.; Haussinger, D.; Baker, D.; Ward, T. R., *J. Am. Chem. Soc.* **2015**, *137*, 10414-10419.
3. Yang, H.; Srivastava, P.; Zhang, C.; Lewis, J. C., *ChemBioChem* **2014**, *15*, 223-227.
4. Song, W. J.; Tezcan, F. A., *Science* **2014**, *346*, 1525-1528.
5. Bos, J.; Browne, W. R.; Driessen, A. J. M.; Roelfes, G., *J. Am. Chem. Soc.* **2015**, *137*, 9796-9799.
6. Srivastava, P.; Yang, H.; Ellis-Guardiola, K.; Lewis, J. C., *Nat. Commun.* **2015**, *6*, 7789.
7. Bos, J.; Roelfes, G., *Curr. Opin. Chem. Biol.* **2014**, *19*, 135-143.
8. Collot, J.; Gradinaru, J.; Humbert, N.; Skander, M.; Zocchi, A.; Ward, T. R., *J. Am. Chem. Soc.* **2003**, *125*, 9030-9031.
9. Pordea, A.; Ward, T. R., *Chem. Commun. (Cambridge, U. K.)* **2008**, 4239-4249.
10. Kohler, V.; Wilson, Y. M.; Durrenberger, M.; Ghislieri, D.; Churakova, E.; Quinto, T.; Knorr, L.; Haussinger, D.; Hollmann, F.; Turner, N. J., *Nat. Chem.* **2013**, *5*, 93-99.
11. Chen, K.; Arnold, F. H., *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 5618-5622.
12. Arnold, F. H., *Acc. Chem. Res.* **1998**, *31*, 125-131.
13. Reetz, M. T., *Angew. Chem., Int. Ed.* **2001**, *40*, 284-310.
14. Hyster, T. K.; Knorr, L.; Ward, T. R.; Rovis, T., *Science* **2012**, *338*, 500-503.
15. Matsuo, T.; Fukumoto, K.; Watanabe, T.; Hayashi, T., *Chem-Asian J* **2011**, *6*, 2491-2499.
16. Dydio, P.; Key, H. M.; Nazarenko, A.; Rha, J. Y.-E.; Seyedkazemi, V.; Clark, D. S.; Hartwig, J. F., *Science* **2016**, *354*, 102-106.
17. Obexer, R.; Godina, A.; Griffiths, A. D.; Garrabou, X.; Mittl, P. R. E.; Baker, D.; Hilvert, D., *Nat. Chem.* **2017**, *9*, 50-56.
18. Ward, T. R., *Acc. Chem. Res.* **2011**, *44*, 47-57.
19. Heinisch, T.; Ward, T. R., *Acc. Chem. Res.* **2016**, *49*, 1711-1721.
20. Romero, P. A.; Arnold, F. H., *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 866-876.
21. Schwizer, F.; Kohler, V.; Durrenberger, M.; Knorr, L.; Ward, T. R., *ACS Catal.* **2013**, *3*, 1752-1755.
22. Kurochkina, N.; Guha, U., *Biophys. Rev.* **2013**, *5*, 29-39.
23. Kramer, M. A.; Wetzel, S. K.; Plückthun, A.; Mittl, P. R. E.; Grütter, M. G., *J. Mol. Biol.* **2010**, *404*, 381-391.
24. Chiu, T. K.; Kubelka, J.; Herbst-Irmer, R.; Eaton, W. A.; Hofrichter, J.; Davies, D. R., *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 7517-7522.
25. Eiben, C. B.; Siegel, J. B.; Bale, J. B.; Shen, B. W.; Foldit Players; Stoddard, B. L.; Popovic, Z.; Baker, D., *Nat. Biotechnol.* **2012**, *30*, 190-194.
26. Coquille, S.; Filipovska, A.; Chia, T.; Rajappa, L.; Lingford, J. P.; Razif, M. F. M.; Thore, S.; Rackham, O., *Nat. Commun.* **2014**, *5*, 5729.
27. Sano, T.; Cantor, C. R., *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 142-146.
28. Le Trong, I.; Humbert, N.; Ward, T. R.; Stenkamp, R. E., *J. Mol. Biol.* **2006**, *356*, 738-745.
29. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., *Nucleic Acids Res.* **2000**, *28*, 235-242.
30. Wood, C. W.; Heal, J. W.; Thomson, A. R.; Bartlett, G. J.; Ibarra, A. A.; Brady, R. L.; Sessions, R. B.; Woolfson, D. N., *Bioinformatics* **2017**, *33*, 3043-3050.

31. Touw, W. G.; Baakman, C.; Black, J.; te Beek, T. A. H.; Krieger, E.; Joosten, R. P.; Vriend, G., *Nucleic Acids Res.* **2015**, *43*, D364-D368.
32. Kabsch, W.; Sander, C., *Biopolymers* **1983**, *22*, 2577-2637
33. Schwizer, F.; Okamoto, Y.; Heinisch, T.; Gu, Y.; Pellizzoni, M. M.; Lebrun, V.; Reuter, R.; Köhler, V.; Lewis, J. C.; Ward, T. R., *Chem. Rev.* **2017**, in press, doi:10.1021/acs.chemrev.7b00014.
34. Heinisch, T.; Langowska, K.; Tanner, P.; Reymond, J.-L.; Meier, W.; Palivan, C.; Ward, T. R., *ChemCatChem* **2013**, *5*, 720-723.
35. Studier, F. W., *Protein Expression Purif.* **2005**, *41*, 207-234.
36. Durrenberger, M.; Heinisch, T.; Wilson, Y. M.; Rossel, T.; Nogueira, E.; Knorr, L.; Zimbron, M. J.; Pierron, J.; Mutschler, A.; Kersten, K.; Schirmer, T.; Ward, T. R., *Angew. Chem., Int. Ed.* **2011**, *50*, 3026-3029.
37. Pellizzoni, M. M.; Gandolfi, R.; Fusè, M.; Facchetti, G.; Contini, A.; Isabella, R., *ChemCatChem* **2016**, *8*, 1665-1670.
38. Jeschek, M.; Reuter, R.; Heinisch, T.; Trindler, C.; Klehr, J.; Panke, S.; Ward, T. R., *Nature* **2016**, *537*, 661-665.
39. Cotellet, Y.; Benz, S.; Avestro, A.-J.; Ward, T. R.; Sakai, N.; Matile, S., *Angew. Chem., Int. Ed.* **2016**, *55*, 4275-4279.
40. López-Andarias, J.; Frontera, A.; Matile, S., *J. Am. Chem. Soc.* **2017**, *139*, 13296-13299.
41. Cotellet, Y.; Lebrun, V.; Sakai, N.; Ward, T. R.; Matile, S., *ACS Cent. Sci.* **2016**, *2*, 388-393.
42. Wang, C.; Matile, S., *Chem. Eur. J.* **2017**, *23*, 11955-11960.

## For Table of Contents Only

