

Functionalized Artificial Bidomain Proteins Based on an α -Solenoid Protein Repeat Scaffold: A New Class of Artificial Diels–Alderases

Thibault Di Meo,^{†,‡,||} Kalani Kariyawasam,^{†,||} Wadih Ghattas,[†] Marie Valerio-Lepiniec,[‡] Giuseppe Sciortino,[§] Jean-Didier Maréchal,[§] Philippe Minard,[‡] Jean-Pierre Mahy,^{*,†} Agathe Urvoas,^{*,‡} and Rémy Ricoux^{*,†}

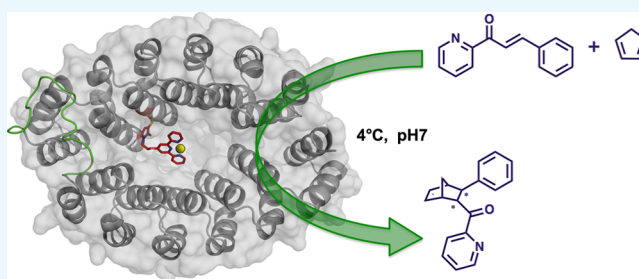
[†]Institut de Chimie Moléculaire et des Matériaux d'Orsay, UMR 8182 CNRS, Laboratoire de Chimie Bioorganique et Bioinorganique, Bât. 420, Univ. Paris-Sud, Université Paris-Saclay, 91405 Orsay Cedex, France

[‡]Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ. Paris-Sud, Université Paris-Saclay, 91198 Gif-sur-Yvette Cedex, France

[§]Departament de Química, Universitat Autònoma de Barcelona, Edifici C.n., 08193 Cerdanyola del Vallés, Barcelona, Spain

S Supporting Information

ABSTRACT: α Rep is a family of entirely artificial repeat proteins. Within the previously described α Rep library, some variants are homodimers displaying interdomain cavities. Taking advantage of these properties, one of these homodimers called α Rep A3 was converted into entirely artificial single chain bidomain metalloenzymes. A non-mutated A3 domain was covalently linked with an A3' domain bearing a unique cysteine on a chosen mutated position (F119C or Y26C). This single mutation ensured the covalent coupling of a 1:1 copper(II)/phenanthroline or copper(II)/terpyridine complex as a catalytic center within the interdomain cavity which was maintained large enough to accommodate two substrates of the Diels–Alder (D–A) reaction. This allowed us to obtain four new artificial Diels–Alderases that were fully characterized by matrix-assisted laser desorption ionization time-of-flight mass spectrometry, UV–vis spectroscopy, and size exclusion chromatography analyses and were then further used for the catalysis of the D–A reaction. They were found to be able to catalyze the enantioselective D–A reaction of azachalcone with cyclopentadiene with up to 38% yield and 52% enantiomeric excess, which validates the proposed strategy. Moreover, the data were rationalized with a computational strategy suggesting the key factors of the selectivity. These results suggest that artificial metalloenzymes based on bidomain A3_A3 proteins modified with nitrogen donor ligands may be suitable for further catalyst optimization and may constitute valuable tools toward more efficient and selective artificial biocatalysts.



INTRODUCTION

Artificial metalloenzymes have emerged as a powerful new approach to expand the repertoire of chemical reactions performed by enzymes.¹ Engineering such new biohybrid catalysts requires a stable host protein displaying a cavity wide enough to accommodate a metal complex responsible for the catalytic process and accessible for the substrates of the catalyzed reaction. α Rep are artificial proteins especially well adapted to the development of a new class of hybrid catalysts. These highly thermostable proteins based on an idealized α -helical repeated motif^{2,3} are devoid of cysteine residues and present a hypervariable surface that can be evolved to create new protein–protein interactions.^{2,4} Recently, described α Rep-based biohybrid catalysts were designed using a homodimeric α Rep named α Rep A3 as protein scaffold. These new artificial Diels–Alderases were obtained by the covalent grafting of a copper(II)–phenanthroline moiety into cysteine-mutated α Rep A3 variants. Positions 26 and 119 were mutated into cysteine and then used for the covalent coupling of the

phenanthroline moiety. The choice of these two positions was based on the crystal structure of α Rep A3.³ For these two positions, the side chains point into the inter-subunit crevice quite deep in the crevice and these positions are among a set of positions highly tolerant to side chain substitutions. In addition, models of two resulting biohybrids, α Rep A3F119C-phenanthroline (A3-F119C-Phen) and α Rep A3Y26C-phenanthroline (A3-Y26C-Phen), suggested that the dimer could accommodate two phenanthroline ligands in the crevice. Finally, the distance between the two β -carbon atoms of symmetrical Tyr26 residues was longer than the distance between the two β -carbon atoms of symmetrical Phe119 residues, suggesting that the resulting biohybrids would present different catalytic properties. The best resulting biohybrid A3F119C-Phen–Cu(II) was found to be able to

Received: December 9, 2018

Accepted: February 13, 2019

Published: February 28, 2019

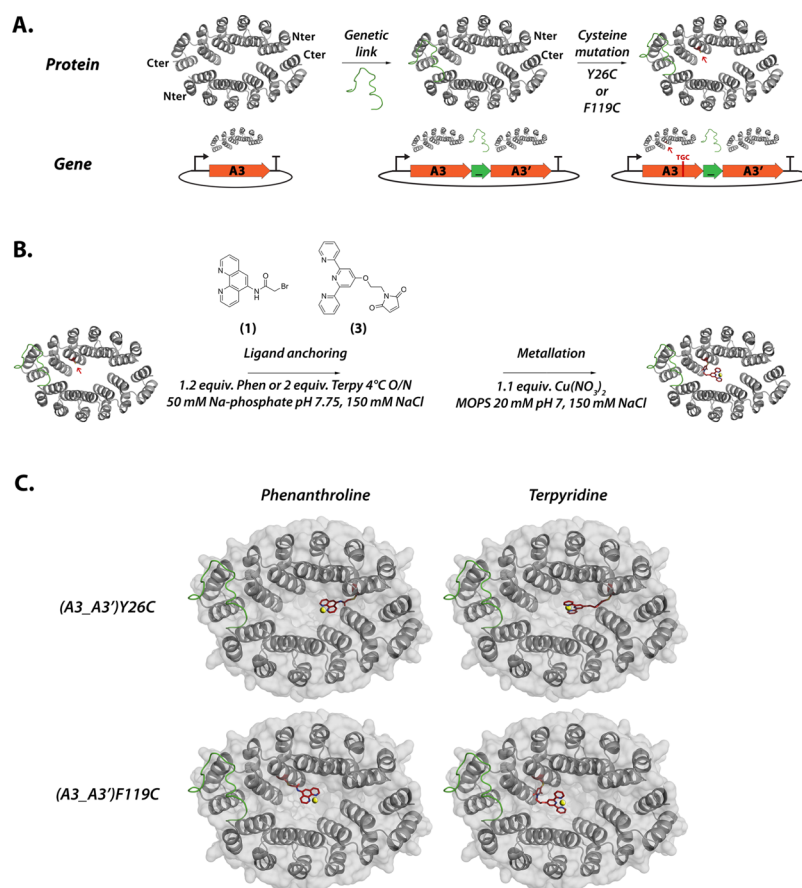


Figure 1. Successive steps for the preparation of biohybrids based on the (A3_A3')F119C and (A3_A3')Y26C variants functionalized with -Phen- and -Terpy-Cu(II) complexes: (A) genetic constructions encoding the bidomain A3_A3' proteins obtained by site-directed cysteine mutations F119C or Y26C. The resulting plasmids were transformed in *E. coli* and the two proteins (A3_A3')F119C and (A3_A3')Y26C were produced and purified. (B) Covalent ligand coupling of phenanthroline or terpyridine functionalized with a maleimide moiety followed by a Cu(II) metallation step. (C) Models for the four resulting biohybrids based on the crystallographic structure of A3 homodimer (3LTM).

enantioselectively catalyze the Diels–Alder cycloaddition of azachalcone with cyclopentadiene with up to 62% enantiomeric excess (ee) but with only 11% yield. Such a low yield could be explained by the presence, as the catalytic species, of the 1:2 copper(II) bis-phenanthroline complex buried in the crevice formed by the protein dimer, which could block the access of the two substrates to the catalytic pocket.³ The formation of such a species was due to the close proximity of the two phenanthroline ligands covalently attached to the two Cys119 residues facing each other in the symmetric dimer. Then, this suggested developing a strategy to anchor only one copper(II) complex as the catalytic center in the catalytic pocket formed by the dimer's interface. In addition, such an approach should maintain a large catalytic pocket in order to accommodate both D–A substrates and thus improve the catalytic properties of the biohybrid catalysts. Such an asymmetric biohybrid cannot be obtained from a homodimeric protein and therefore needs to be created from one protein scaffold containing two different domains linked together in order to obtain an asymmetric bidomain protein. The metal complex can then be grafted on a single cysteine located on one of the two domains while the second domain contains no cysteine residue. We recently described the construction of the single chain bidomain protein A3_A3 in which two A3 domains were encoded by different genetic sequences and linked by a flexible (SGGGG)₂ peptide.⁵

Herein, we report the successful production of single-chain bidomain proteins in which a cysteine-free domain was linked with a domain bearing either the F119C mutation [(A3_A3')F119C] or the Y26C mutation [(A3_A3')Y26C]. Further chemical coupling of either a phenanthroline ligand (Phen) or a bulkier terpyridine ligand (Terpy) followed by insertion of a copper(II) ion led to four new artificial metalloenzymes ((A3_A3')F119C-Phen-Cu(II), (A3_A3')F119C-Terpy-Cu(II), (A3_A3')Y26C-Phen-Cu(II), and (A3_A3')Y26C-Terpy-Cu(II)) that were characterized by matrix-assisted laser desorption ionization time-of-flight (MALDI-ToF) mass spectrometry (MS) and size exclusion chromatography. This new class of artificial metalloenzymes was found to be able to catalyze the enantioselective Diels–Alder (D–A) reaction of azachalcone with cyclopentadiene presenting up to 38% yield and ee up to 52%, which validates the proposed strategy. Molecular modeling allowed the identification of the preferential orientations of the two copper complexes in the binding pocket. The computation allowed suggesting the key recognition factors of the stereo determining step of the reaction.

RESULTS AND DISCUSSION

Engineering of the Host Protein. As the first α Rep-based biohybrid design was based on the α Rep A3 homodimer, the introduction of a single cysteine in the polypeptide chain

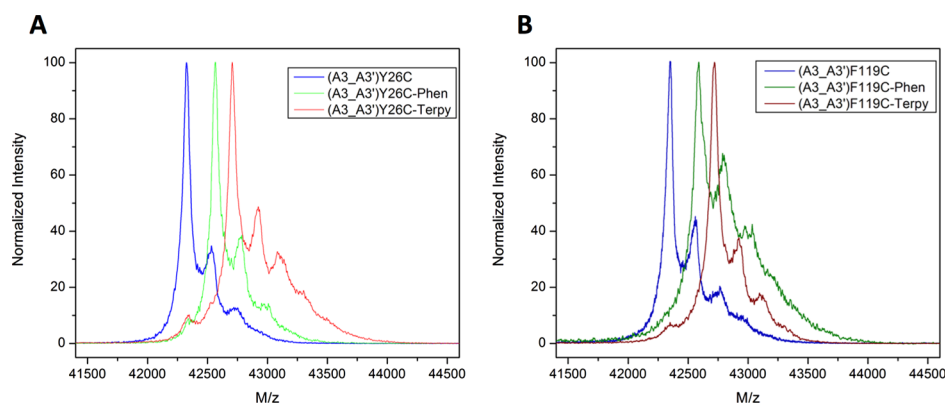


Figure 2. MALDI-ToF spectra of the (A3_A3')Y26C (A) and (A3_A3')F119C (B) protein variants measured before (blue) and after functionalization with terpyridine (red) and phenanthroline (green).

resulted in the presence of two cysteine residues within the dimer's cavity and therefore necessarily ended up with a symmetrical coupling of two chemical ligands. In order to explore the effect of a single metal complex anchored in the cavity of this protein, we needed to mutate one residue of only one domain of the dimer into cysteine. The differentiation of the two domains of the dimer can be achieved by mutation if they were expressed as in a bidomain monomeric protein. We recently described the genetic construction of a plasmid encoding such an A3_A3 bidomain monomeric protein, in which a synthetic nonrepeated gene encoded the first domain and the repeated sequence contained in the α Rep library encoded the second domain. The two domains are covalently linked on a single polypeptide chain by a flexible linker sequence. The X-ray structure of this nonmutated bidomain A3_A3 protein (PDB ID 6FT5) showed that the linker did not affect the global structure of the monomer and functional studies indicated that the linker reinforced the closed form of the bidomain protein.⁵ Taking advantage of the nonrepeated sequence of the first domain, mutations could selectively be introduced in this domain. In order to compare the activities of the resulting biohybrids with the previously described dimer's activities, the same positions as in the dimer were chosen for cysteine mutation (F119 and Y26). The corresponding synthetic genes obtained by site directed mutagenesis were sub-cloned into a modified expression vector, which resulted in plasmids encoding the two bidomain variants (A3_A3')F119C and (A3_A3')Y26C. The proteins were produced in *Escherichia coli* and purified as described and yields were equivalent to the ones obtained with the dimers (between 50 and 70 mg of pure protein per liter of culture). Analytic size exclusion chromatography profiles were equivalent to that of the A3 dimer and to that of the nonmutated A3_A3 bidomain protein, indicating that the mutations did not affect the global protein conformation.

Bioconjugation. For this study, we selected two robust copper(II) ligands with different nitrogen donor motifs that is, phenanthroline and terpyridine, which are known to bind copper(II) with strong affinities to give Cu(II) complexes and that are able to catalyze D–A reactions.^{6,7} Derivatives of these ligands had to be prepared to enable their covalent attachment to the (A3_A3')F119C and (A3_A3')Y26C variants. 2-Bromo-*N*-(1,10-phenanthrolin-5-yl)acetamide (**1**) and *N*-[2-([2,2':6',2''-terpyridin]-4'-yloxy)ethyl]maleimide (**3**) were prepared following already described procedures (see [Experimental Section](#)) as they bear reactive bromoacetyl group and

maleimido group, respectively, which could react with the thiol function of the cysteinyl residues of the mutant proteins ([Figure 1](#)). The reactivity and accessibility of the thiol function of the cysteinyl residues was then assessed, thanks to the time-resolved Ellman's assay using sulfhydryl reagent 5,5-dithio-bis-(2-nitrobenzoic acid) (DTNB) that reacts with sulfhydryl groups, releasing 2-nitro-5-thiobenzoate (TNB[−]) which displays a maximum absorption at 412 nm.⁸ First, for each one of the two variants, the labeled cysteine concentration at the end of the reaction was equal to the protein concentration. This indicated that all cysteine residues were reactive and did not form any disulfide bridge even after several weeks' storage without any reducing agent. Second, the kinetic of the reaction was different for the two variants: (A3_A3')Y26C showed very fast labeling kinetics and led rapidly to a stable maximum absorbance at 412 nm: ($t_{1/2} = 30$ s), while (A3_A3')F119C reacted more slowly with DTNB and the absorbance at 412 nm reached a plateau value only after several minutes: ($t_{1/2} = 3$ min 30 s).

This denoted in (A3_A3')F119C either a sterically hindered environment or a shift in the pK_a of Cys119. Thus, the thiol function appeared to be more accessible and more reactive in (A3_A3')Y26C than in (A3_A3')F119C. It is noteworthy that, in both cases, those $t_{1/2}$ values were longer than those measured for the symmetrical (A3Y26C)₂ and (A3F119C)₂ containing no linker (resp. 30 s vs instantaneous and 3 min 30 s vs 1 min 40 s).³ This indicated that in the single chain (A3_A3')Y26C and (A3_A3')F119C proteins the linker peptide stabilizes the closed form of the binding cavity. This result is in accord with by the functional studies reported for the nonmutated A3_A3, indicating that the interdomain interactions are reinforced in the bidomain as compared to the A3 homo-dimer.⁵

Protein Functionalization. The two variants (A3_A3')Y26C and (A3_A3')F119C were then functionalized by incubation with 1.2 equiv of 2-bromo-*N*-(1,10-phenanthrolin-5-yl)acetamide hydrobromide or 2 equiv of *N*-[2-([2,2':6',2''-terpyridine]-4-yloxy)ethyl]maleimide overnight at 4 °C, resulting in the formation of the cysteine–phenanthroline (Phen) or cysteine–terpyridine (Terpy) labeled proteins in their *apo* form: (A3_A3')Y26C-Phen, (A3_A3')F119C-Phen, (A3_A3')Y26C-Terpy, and (A3_A3')F119C-Terpy. The complete functionalization was confirmed by Ellman's assay, with a complete loss of thiol reactivity (data not shown).

MALDI-ToF MS analysis was first used to characterize the biohybrids ([Figure 2](#)). Peaks at $42\,564 \pm 5$ and $42\,706 \pm 5$ Da

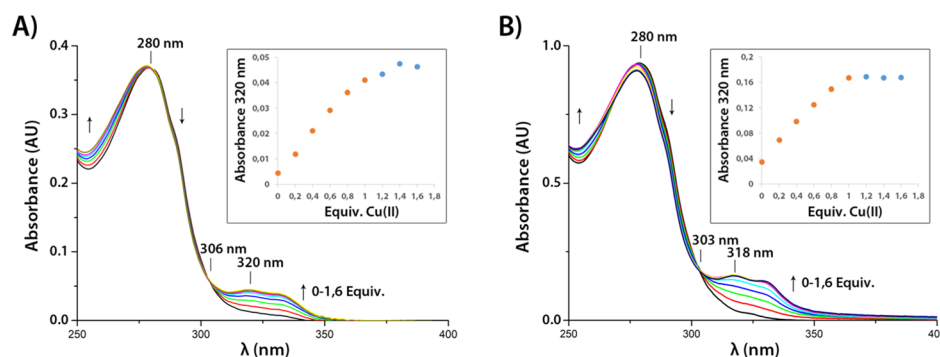


Figure 3. Metallation of the (A3_A3')Y26C-Terpy and (A3_A3')F119C-Terpy biohybrids with Cu(II). (A) Evolution of the UV–visible spectrum of (A3_A3')Y26C-Terpy after the progressive addition of up to 1.6 equiv $\text{Cu}(\text{NO}_3)_2$ to a $5.6 \mu\text{M}$ solution of (A3_A3')Y26C-Terpy in 20 mM MOPS 150 mM NaCl pH 7 at 20°C . Inset: evolution of the absorbance at 320 nm as a function of the number of equiv. of $\text{Cu}(\text{NO}_3)_2$ /(A3_A3')Y26C-Terpy. (B) Evolution of the UV–visible spectrum of (A3_A3')F119C-Terpy after the progressive addition of up to 1.6 equiv $\text{Cu}(\text{NO}_3)_2$ to a $13.5 \mu\text{M}$ solution of (A3_A3')Y26C-Terpy in 20 mM MOPS 150 mM NaCl pH 7 at 20°C . Inset: evolution of the absorbance at 318 nm as a function of the number of equiv of $\text{Cu}(\text{NO}_3)_2$ /(A3_A3')F119C-Terpy.

were observed respectively for (A3_A3')Y26C-Phen and (A3_A3')Y26C-Terpy that were in agreement with the covalent anchoring of one phenanthroline and one terpyridine per protein (calculated mass 42 565 and 42 702 Da, respectively) (Figure 2A). Peaks at $42\,585 \pm 5$ and $42\,720 \pm 5$ Da were also observed in the mass spectra of (A3_A3')F119C-Phen and (A3_A3')F119C-Terpy, respectively (Figure 2B), that were in agreement with the covalent anchoring of one phenanthroline and one terpyridine ligand per protein (calculated mass 42 581 and 42 718 Da, respectively).

When the functionalization experiment was performed on the A3_A3 protein bearing no cysteine residues, the MALDI-ToF MS analysis showed only a peak at 42 346 Da corresponding to the noncoupled protein, which indicated that the coupling reaction only occurred in the presence of cysteine residues in the variants and that no other reactive group was modified (data not shown).

Insertion of Copper(II) into the (A3_A3')Y26C and (A3_A3')F119C Variants Functionalized with Phen and -Terpy. The (A3_A3')Y26C-Phen–Cu(II) and (A3_A3')F119C-Phen–Cu(II) biohybrids did not present any characteristic absorbance in the 300–400 nm range that could allow us to follow the binding of the copper(II) ion to the ligand. On the contrary, for (A3_A3')Y26C-Terpy–Cu(II) and (A3_A3')F119C-Terpy–Cu(II) biohybrids, the insertion of Cu(II) into the terpyridine ligand was accompanied by the appearance of two bands at 320 and 330 nm assigned to the $\pi \rightarrow \pi^*$ transition of the Terpy–Cu(II) complex (Figure 3). It was therefore possible to carry out a titration of the (A3_A3')Y26C-Terpy and (A3_A3')F119C-Terpy biohybrids by increasing amounts of copper(II) nitrate by following the absorbance at 320 nm as a function of the $\text{Cu}(\text{NO}_3)_2$ /(A3_A3')Y26C-Terpy and (A3_A3')F119C-Terpy ratio.

In a typical experiment, the effect of the successive additions of 0.2 equiv of $\text{Cu}(\text{NO}_3)_2$, from 0 to 1.6 equiv, to a $5.6 \mu\text{M}$ solution of (A3_A3')Y26C-Terpy in 20 mM MOPS 150 mM NaCl pH 7, was followed by UV–visible spectroscopy (Figure 3A). This allowed us to follow the gradual appearance of the bands at 320 and 330 nm characteristic of the (A3_A3')Y26C-Terpy–Cu(II) complex and the formation of an isosbestic point at 306 nm. When plotting the absorbance at 320 nm as a function of the number of equiv of Cu(II) added to (A3_A3')Y26C-Terpy (Figure 3A, inset), a typical saturation curve was obtained with a linear increase, till to 1 equiv of

$\text{Cu}(\text{NO}_3)_2$ added (red dots), of the absorbance at 320 nm which then remains about constant for higher concentrations of $\text{Cu}(\text{NO}_3)_2$ (blue dots). This is clearly the sign of the formation, with a 1:1 stoichiometry, of a (A3_A3')Y26C-Terpy–Cu(II) complex in which the Terpy ligand displayed a high affinity for the copper(II) ion. The same experiment was performed with (A3_A3')F119C-Terpy at a concentration of $13.5 \mu\text{M}$ and similar results were obtained (Figure 3B), the isosbestic point being at 303 nm and the maximum absorbance wavelength of the bands being 318 and 330 nm in this case. Moreover, the calculation of molar extinction coefficient of the complex at 318 nm resulted in a value of $6500 \text{ M}^{-1}\cdot\text{cm}^{-1}$ for (A3_A3')Y26C-Terpy–Cu(II) and $9500 \text{ M}^{-1}\cdot\text{cm}^{-1}$ for (A3_A3')F119C-Terpy–Cu(II). These shifts can be caused by either the small differences between the spectral properties of the two variants, or by a different environment of the terpyridine moiety inside the (A3_A3')Y26C and (A3_A3')F119C pocket.

It is noteworthy that when the solution was subjected to chromatography on a column of molecular sieves, which removed the excess of copper nitrate, the A3sA3Y26Terpy–Cu(II) complex was recovered unchanged, as shown by its UV–visible spectrum. This confirmed the high affinity of (A3_A3')Y26C-Terpy for Cu(II). However, it was possible to remove Cu(II) from the Terpy ligand by incubating the purified complex with a large excess of ethylenediaminetetraacetic acid (EDTA) (5 mM).

Consequently, for further use as catalysts for the D–A reaction, 100 μM solutions of (A3_A3')F119C-Phen–Cu(II), (A3_A3')Y26C-Phen–Cu(II), (A3_A3')F119C-Terpy–Cu(II), or (A3_A3')Y26C-Terpy–Cu(II) in buffer A were prepared by adding 1 equiv of copper nitrate (10 mM in water) to the corresponding biohybrids solutions, as described in the Experimental Section. The final solutions were chromatographed on molecular sieves prior to use in order to remove any excess of Cu(II) that could bind nonspecifically to the protein.

Analytic Size–Exclusion Profiles of the Biohybrids. In order to ensure that phenanthroline and terpyridine grafting on the (A3_A3')Y26C and (A3_A3')F119C variants did not disrupt the closed conformation of A3_A3 in solution, the oligomeric states of the biohybrids copper complexes of (A3_A3')F119C-Phen, (A3_A3')Y26C-Phen, (A3_A3')-

F119C-Terpy, and (A3_A3')Y26C-Terpy were investigated by gel filtration (Figure 4).

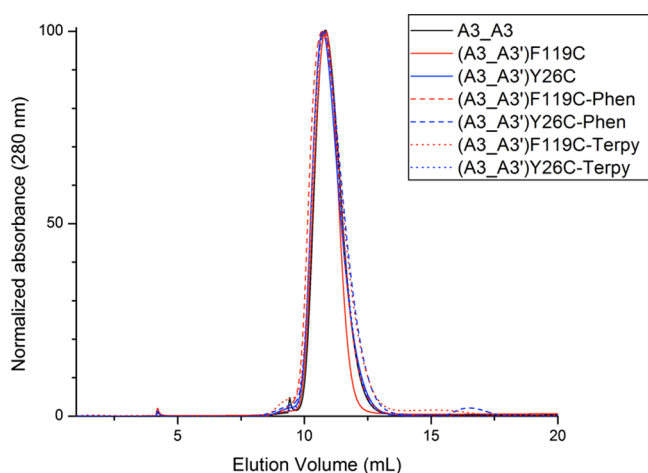


Figure 4. Size-exclusion profiles of the biohybrids. Solutions containing sample proteins were injected on Superdex 75 10/300 GL column equilibrated in buffer A and elution profiles were recorded using the absorbance at 280 nm. For all the samples, an excess of copper(II) per protein monomer was pre-incubated before injection on the column. Normalized elution profiles are shown for A3_A3 (black), (A3_A3')F119C (red), (A3_A3')Y26C (blue), and the corresponding biohybrids (A3_A3')F119C-Phen-Cu(II) (red, dashed line), (A3_A3')F119C-Terpy-Cu(II) (red, dotted line), (A3_A3')Y26C-Phen-Cu(II) (blue, dashed line) and (A3_A3')Y26C-Terpy-Cu(II) (blue, dotted line).

First, the elution volumes of the two variants (A3_A3')-F119C (10.84 mL) and (A3_A3')Y26C (10.72 mL) were very close to the elution volume of wild type A3_A3 (10.87 mL) in the presence of Cu(II), indicating that the cysteine mutations did not affect the monomeric structure and overall closed conformation of the single chain bidomain protein. Then, the elution volumes observed for the four biohybrids (10.70 mL) for (A3_A3')F119C-Phen-Cu(II), 10.77 mL for (A3_A3')Y26C-Phen-Cu(II), 10.76 mL for (A3_A3')F119C-Terpy-Cu(II), and 10.78 mL for (A3_A3')Y26C-Terpy-Cu(II) remained very similar reflecting a very close hydrodynamic radius value for all the proteins. This result showed that the closed form of the A3_A3 protein was retained for the four biohybrids and that the covalent attachment of both Terpy-Cu(II) and Phen-Cu(II) complexes to the (A3_A3')Y26C and (A3_A3')F119C variants did not alter the overall conformation of the protein.

Catalysis of a Diels–Alder Reaction. The Diels–Alder (D–A) cycloaddition is a very useful carbon–carbon bond-forming reaction for the synthesis of organic building blocks, and it has been widely studied.^{9,10} Very few enzymes are known to catalyze this reaction,¹¹ therefore artificial enzymes

with this activity present a high potential for synthetic applications.¹² As a consequence, in the last decade, many artificial Diels–Alders have been reported for the catalysis of the benchmark reaction of azachalcone with cyclopentadiene and showed variable endo/exo selectivities and enantioselectivities.^{7,13–15}

The reaction of 2-azachalcone (**4**) with cyclopentadiene (**5**) (Scheme 1) was thus investigated in the presence of either the unmodified A3_A3 protein or the (A3_A3') cysteine mutant-derived biohybrids, in the presence or not of copper(II)–phenanthroline or –terpyridine complexes. It is noteworthy that this reaction can result in up to four isomer products: two major endo enantiomers (**6a** and **6b**) and two minor exo enantiomers (**6c** and **6d**, Scheme 1). Consequently, in all the cases, as the exo products were negligible versus endo products, ee values were calculated for the endo products. In all the cases where an enantiomeric excess was observed, the major enantiomer was the 1S,2R,3R,4R endo product **6a** with respect to the 1R,2S,3S,4S endo product **6b** as determined by circular dichroism analysis.²³

The D–A reactions were performed under the following conditions: 34 mM **5** and 1 mM **4** were allowed to react in buffer A, and no co-solvent added apart from the substrate solution (1% acetonitrile) at 4 °C, in the presence of 30 μM of catalyst (Table 1).

Table 1. Compared ee, Yields, and Endo/Exo Ratio Measured after the D–A Reaction of Cyclopentadiene **5** with 2-Azachalcone **4** in MOPS Buffer Catalyzed by Either A3_A3, (A3_A3')Y26C-Phen, (A3_A3')Y26C-Terpy, (A3_A3')F119C-Phen, or (A3_A3')F119C-Terpy Biohybrids in the Presence of Copper(II)nitrate^a

catalyst	ee (%)	yield (%)
Cu(NO ₃) ₂	0 ± 1	38 ± 2
A3_A3	15 ± 3	7 ± 1
A3_A3 + Cu	3 ± 2	22 ± 3
A3_A3 + Phen-Cu	31 ± 1	9 ± 1
A3_A3 + Terpy-Cu	19 ± 1	7 ± 1
(A3_A3')Y26C-Phen	12 ± 1	8 ± 2
(A3_A3')Y26C-Terpy	16 ± 1	9 ± 2
(A3_A3')F119C-Phen	ND ^b	4 ± 1
(A3_A3')F119C-Terpy	16 ± 1	6 ± 1
(A3_A3')Y26C-Phen-Cu	22 ± 1	21 ± 2
(A3_A3')Y26C-Terpy-Cu	14 ± 1	16 ± 2
(A3_A3')F119C-Phen-Cu	5 ± 1	38 ± 2
(A3_A3')F119C-Terpy-Cu	52 ± 1	15 ± 1

^aIn all experiments, 30 μM monomer protein was used after incubation with 1 equiv of copper(II)nitrate. ^bee value could not be determined in this case due to the low yield.

Under these conditions, in the absence of copper(II) nitrate, very low yields were obtained, endo products **6a–b** being

Scheme 1. Diels–Alder (D–A) Reaction

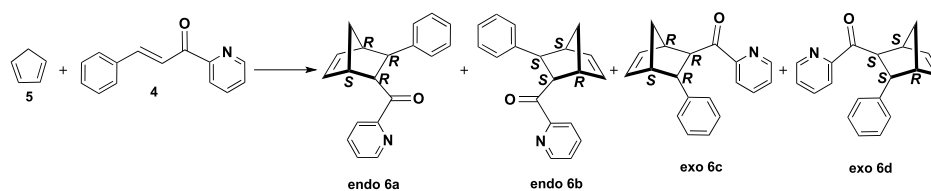


Table 2. Selected GoldScore Values for the Docking of Metal Free Cofactor, Metalated Cofactor + Substrate, and Metalated Cofactor + Endo-6a and 6b Products^a

(A3_A3') variant	isomer products	yield ^{exp}	F_{\max}	F_{\max} (Cu + 4)	ee ^{exp} (%)	F_{\max} (Cu + 6)	ΔF_{\max}
F119C-Phen	6a	38 ± 2	26.6 (9/50, I)	30.0 (45/50, II)	5	55.1	3.8
	6b					51.3	
Y26C-Phen	6a	21 ± 1	26.6 (6/50, IV)	29.8 (46/50, I)	22	33.6	5.6
	6b					28.0	
F119C-Terpy	6a	15 ± 1	40.3 (33/50, III)	42.0 (23/50, I)	52	61.7	7.2
	6b					54.5	
Y26C-Terpy	6a	16 ± 2	41.0 (21/50, II)	37.9 (1/50, IIX)	14	54.1	5.2
	6b					48.9	

^aThe data are compared with relevant experimental yield and ee data. In parenthesis, the position in ranking and the population of the specific cluster are reported.

obtained in about 7 ± 1% yield, even in the presence of A3_A3 with an enantiomeric excess of 15 ± 3% (Table 1). Copper(II) nitrate alone was found to be able to catalyze the D–A reaction, a 38 ± 2% yield in products **6** being obtained in the presence of 30 μM Cu(NO₃)₂, but no enantiomeric excess was observed. When both A3_A3 and copper(II) nitrate were associated, a 22 ± 1% yield was obtained, but with a negligible enantiomeric excess of about 3 ± 2%. When Phen–Cu and Terpy–Cu were noncovalently inserted into A3_A3, to generate the (A3_A3)-Phen–Cu and (A3_A3)-Terpy–Cu biohybrids following the “host–guest” strategy, low yields in endo products were formed (resp. 7 and 9 ± 1%) but higher ee of respectively 31 ± 1 and 19 ± 1% were measured.

When the reactions were run in the presence of either the four biohybrids, (A3_A3')Y26C-Phen or -Terpy, and (A3_A3')F119C-Phen or -Terpy, but in the absence of copper(II), rather low yields ranging from 6 ± 1 to 9 ± 1% were obtained.

When the (A3_A3')Y26C derived biohybrids were used in the presence of copper(II), quite similar results were observed, (A3_A3')Y26C-Phen–Cu leading both to slightly higher yield (21 ± 2%) and enantiomeric excess (22 ± 1%) than (A3_A3')Y26C-Terpy–Cu (16 ± 2% yield and 14 ± 1% ee).

When the reactions were catalyzed by (A3_A3')F119C derived biohybrids, two different results were observed. First, when (A3_A3')F119C-Phen–Cu(II) was used as catalyst, a higher yield in product **6** (38 ± 2%) was obtained but almost no enantiomeric excess was detected (5 ± 1%). On the contrary, when the Phen ligand was replaced by the more bulky ligand terpyridine (A3_A3')F119C-Terpy–Cu(II), the enantiomeric excess increased to reach 52 ± 1% but the yield in product **6** decreased to 15 ± 1%.

Molecular Modeling. A computational approach was subsequently performed to rationalize these data. Our strategy consisted in (i) building the two models of the biohybrid variants (A3_A3')Y26C and (A3_A3')F119C; (ii) characterizing throughout density functional theory (DFT) calculations on cluster models the geometries of intermediates and products along the copper mediated D–A reaction; and (iii) inserting those geometries into the variant models by optimized protein–ligand docking. The complementarity of the catalytically relevant structures into the context of the catalysis was then discussed. More concretely, the modeling consisted in building the biohybrid structures of the metal free cofactors covalently linked to (A3_A3')Y26C and (A3_A3')-F119C following the three different reaction steps: (1) the resting state of the artificial metalloenzymes produced by metalation of the hybrids (docking of Cu(cofactor)(H₂O)₂),

(2) ligand exchange leading to the formation of the co-substrate-bound cofactor [docking of Cu(cofactor)(**4**)], and (3) metal containing cofactor bound to endo-**6a**, -**6b** and exo-**6c**, -**6d** D–A products to model the final step of the reaction [docking of Cu(cofactor)(**6**)]. Because transition states full characterization are not achievable by docking methods, and accurate QM/MM calculations are out of the scope of the present paper, simulation with product-like geometry were expected to provide the main information concerning protein–ligand recognition factors that could define the stereospecificity of the biohybrid enzymes.²⁴ Details on the results and their analysis are reported in the Supporting Information.

When comparing the docking on the different structural models of the reaction taking place with the biohybrids, interesting structural tendencies appear. The calculations on metal-free Phen and Terpy biohybrids show substantial cofactor flexibility in the low energy solutions with the cofactor being able to go from subpockets to different ones for (A3_A3') F119C-Phen and for (A3_A3')F119C-Terpy. The cofactor goes from the subpockets involving residues S147, W143, D186, Q146, Y211', W235', D242', Q246' and W143, Q146, S147, Y26', G29', K30', W50', A54', A58', Q61', Q246' to subpockets involving residues A116, I115, F112, Q34, A116, V88 and W81, W143, W81', Q84', V88', S85', R53', D57', W111', F112', I115', for (A3_A3')F119C-Phen and (A3_A3')F119C-Terpy, respectively. This scattering is though largely reduced upon the loading of the metal and its *aquo* ligands leading to low energy solutions with the cofactors and Cu(II) directed toward the center of the hydrophobic binding pocket (Table S1C). The major interactions for the lower energy solutions are between the water molecules and side chains of S85', D57', R53', and R238' as well as hydrophilic contacts of the cofactor with W143, Q146, W50', Y26', and W143, Q146, W235', Q246', V273' for (A3_A3')F119C-Phen–(H₂O)₂ and (A3_A3')F119C-Terpy–(H₂O)₂, respectively (see Supporting Information). These low energy orientations are consistent with those obtained when the 2-azachalcone substrate binds to the copper showing that the metalated biohybrids present a good pre-organization of the binding site for the first co-substrate. Finally, docking simulations with the D–A products endo-**6a**, -**6b**, exo-**6c**, and -**6d** all show good complementarities with the binding site (more than 50.0 GoldScore units) though at this state we can already observe that endo-**6a** products have better complementarity values than endo-**6b** with at least 5.2 GoldScore fitness units (Table S2C).

In order to better understand the key recognition factors involved in the enantiomeric selectivity, the dockings for the

(A3_A3')F119C-Phen-Cu-**6a** and **6b** (5% ee) and (A3_A3')F119C-Terpy-Cu-**6a** and **6b** (52% ee) systems were repeated considering the flexibility in solution of the side chains in the binding pocket. All the data taken together could be rationalized considering the steric hindrance of the binding sites and the stabilization of the products with a transition state-like geometry (Table 2).

In the case of (A3_A3')F119C-Phen-Cu-(2-azachalcone), it appeared that the Phen ligand and the substrate were surrounded by a wide pocket showing van der Waals contacts with V88, W143, Q84', W99', F112', I115', and W143' (Figure 5a). The relatively open pocket and the limited

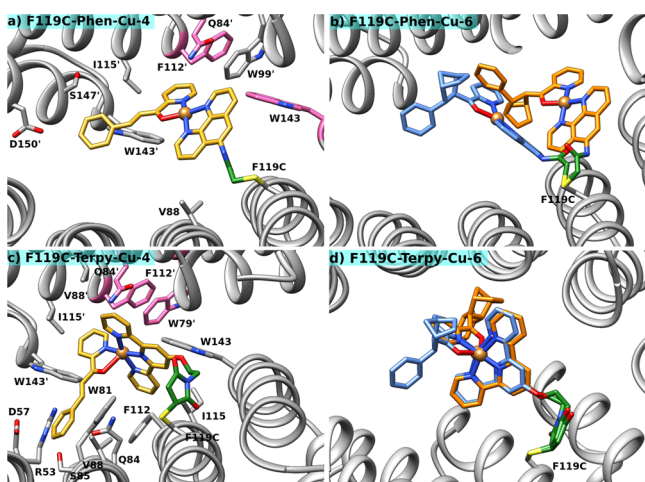


Figure 5. Representation of best docking solutions for the reactants (A3_A3')F119C-Phen-Cu(4), (A3_A3')F119C-Phen-Cu(6) and (A3_A3')F119C-Terpy-Cu(4), (A3_A3')F119C-Terpy-Cu(6) unit showing their environment. The linker is depicted in green and the side chains that change their conformation during binding are highlighted in pink.

number of interacting residues could lead to a moderate control of the relative positioning of both complex and substrates, in fact, for (A3_A3')F119C-Phen-Cu(endo-6), both isomers reach binding modes with similar scoring ($\Delta F_{\max} = 3.8$, Table 2). These considerations could explain both the low ee observed as well as the higher yield in product 6 ($38 \pm$

2%). On the contrary, in the case of (A3_A3')F119C-Terpy-Cu(2-azachalcone), the bulkier maleimide linker and Terpy ligand lead to stronger van der Waals stabilizations of the complex with R53, Q84, S85, V87, W81, F112, I115, W143, W81', Q84', V88', F112', I115', and W143' that hinder its flexibility and the access of the cyclopentadiene to one side of the pocket (Figure 5c). Moreover, the docking analysis of (A3_A3')F119C-Terpy-Cu(endo-6a/6b) shows an equivalent position of the terpyridine implying the specular orientation of the two products that leads to favor the **6a** enantiomer stabilization, in particular the interactions of the substrates with Q84, F112, I115, A116, W143, Q84', V88', F112', I115', F119', and W143' ($\Delta F_{\max} = 7.2$, Figures 5d and S2C). These two observations could explain both the increase in the enantiomeric excess (5–52%) and the decrease in the yield (38–15%).

Concerning the (A3_A3')Y26C variant, the docking analysis shows a significant stabilization of the intermediate Cu-(cofactor)(2-azachalcone) with residues Q61, Q84, V88, A116, I115, F119, W81', Q83', I115', W143', S148' and Q84, V88, F112, A116, W111, I115, F119, W143, Q85', V88', I115', A116', F119' for (A3_A3')Y26C-Phen-Cu(4) and (A3_A3')Y26C-Terpy-Cu(4), respectively. The preferential formation of the **6a** enantiomer is predicted obtaining fitness differences of 5.6 and 5.2 GoldScore units for Phen and Terpy variants, respectively. In analogy with the previous case, the (A3_A3')Y26C-Terpy-Cu(2-azachalcone) shows both a larger pocket and a higher variability in docking solutions. Accordingly, a yield in product **6** of $16 \pm 2\%$ and an ee of $14 \pm 1\%$ were found for (A3_A3')Y26C-Phen-Cu(II) and a $21 \pm 2\%$ yield in product **6** for (A3_A3')Y26C-Phen-Cu(II) with an ee of $22 \pm 1\%$. Both species are stabilized by van der Waals contacts with V88, Q84, Q61, V88', I115', F119', W143', S147', D150' for (A3_A3')Y26C-Phen-Cu(6a) and V88, Q84, W81, F119, W143, W81', Q84', V88', A116', F119', W143' for (A3_A3')Y26C-Terpy-Cu(6a) (Figure S1C).

Taken all together, the computation results suggest possible further improvements to our system by, for example, stabilizing the intermediates with bulkier hydrophobic residues in their direct vicinity or hinder the reactants access to one side of the binding pocket favoring the selective formation of one enantiomer. It is noteworthy that no significant effect of the

Table 3. Compared Yields and ee Measured after the D–A Reaction of Cyclopentadiene 5 with 2-Azachalcone 4 in the Presence of Various Copper(II)-Biohybrid Catalysts Reported in the Literature

biohybrid catalyst protein–ligand ^a	conversion/yield (%)	endo/exo (%)	ee (%)	refs
LmrR_M89C-Phen ^b	93	95:5	97	7
SB-Lys-GTL*/196-Phen ^b	98	93:7	25	17
HSA-phtalocyanin	89	91:9	85	18
NCS-Phen ^b	84	84:16	0	10,16
FhuA-NHC ^c	62	90:10	ND	19
FhuA-Tpy ^d	15	66:34	ND	10
A ₂ A-Phen ^d	28	82:18	19	14
NB-Tpy ^d	22	90:10	ND	6
A3F119C-Phen ^b (A3F119NPH)	11	86:14	62	3
(A3_A3')F119C-Terpy	15	93:7	52	this work
(A3_A3')F119C-Phen	38	92:8	5	

^aNomenclature for the various proteins used: LmrR, *Lactococcal* multidrug resistance regulator; SB-Lys-GTL*/196, Sepabeads Lys196 mutant of *Geobacillus thermocatenulatus* lipase; HSA, human serum albumin; A3, α Rep A3 protein; SCP-2L, sterol carrier protein type 2-like domain; NCS, neocarzinostatin; NB, nitrobindin; FhuA, ferric hydroxamate uptake protein component A. A₂A, A₂ adenosine receptor. ^bPhen, phenanthroline. ^cNHC, N-heterocyclic carbene. ^dTpy, terpyridine.

catalysis on regioselectivity was observed with either of the two catalytic Cu(II) complexes in the A3_A3' mutants as in all the cases the endo/exo ratios were similar ($92 \pm 1:8 \pm 1$).

DISCUSSION

The characteristics of the D–A reaction catalyzed by (A3_A3')F119C-Terpy- and -Phen–Cu(II), yield, endo/exo ratio, and enantiomeric excess, ee, are compared in Table 3 with those reported to date for the D–A reaction between 4 and 5 catalyzed by various artificial metalloenzymes.

Considering the endo/exo ratio, (A3_A3')F119C-Terpy–Cu(II) and (A3_A3')F119C-Phen–Cu(II) (endo/exo = 93/7 and 92/8 resp.) stand among the systems that follow the traditional endo rule with LmrR_M89C-Phen–Cu,⁷ SB-Lys-GTL*/196-Phen–Cu¹⁷ (endo/exo = 95/5 and 93/7, respectively), whereas FhuA-Tpy favored the formation of the exo products (endo/exo = 66/34).¹⁰ Concerning the yields in product 6, the best one that was obtained in this work, 38% with (A3_A3')F119C-Phen–Cu, was lower than those obtained with the most efficient systems LmrR_M89C-Phen–Cu,⁷ SB-Lys-GTL*/196-Phen–Cu,¹⁷ and HSA-phtalocyanin–Cu¹⁸ that ranged between 89 and 98%, but were higher than those reported for NB-Tpy–Cu⁶ and FhuA-Tpy–Cu¹⁹ (22 and 15%, respectively). Finally, the best ee obtained, 52% with (A3_A3')F119C-Terpy–Cu, was also lower than those reported with LmrR_M89C-Phen (ee = 97%)⁷ and HSA-phtalocyanin (ee = 85%)¹⁸ but higher than that reported with SB-Lys-GTL*/196-Phen (ee = 25%).¹⁷ We recently published the production of a NCS–phenanthroline–Cu(II) artificial metalloenzyme, by the non-covalent insertion of a testosterone–phenanthroline–Cu(II) conjugate into the binding site of NCS following the “Trojan Horse” strategy.¹⁶ This biohybrid was found able to catalyze the D–A reaction with a very good yield, 84%, but with no enantiomeric excess, which could be explained both by the wide binding site of NCS open to the solvent, with no control of the relative positioning of the two D–A substrates, and the high flexibility of the catalytic Phen–Cu(II) moiety due to its noncovalent anchoring to the protein. We then reported a new artificial Diels–Alderase A3F119NPH–Cu(II) that was obtained by covalent anchoring of a copper(II)–phenanthroline moiety into the F119C mutant of an artificial protein dimer based on a thermostable α -helical repeated motif.³ In this case, a good enantioselectivity (62% ee) was obtained but with the yield in product 6 was low (11%), which could be explained by the presence of a copper(II)bisphenanthroline complex, due to the close proximity of the two phenanthroline ligands covalently attached to the two Cys119 residues facing each other in the symmetrical dimer, which blocked the access of the two substrates to the catalytic center.³ Consequently, in order to avoid this major drawback, we proposed to associate in a single chain protein a single cysteine mutated A3 domain with another nonmutated A3 domain, in order to ensure to have a 1:1 copper(II)–phenanthroline complex as catalytic center and to enlarge the catalytic pocket in order to accommodate both D–A substrates. The results presented in this paper show that this strategy was successful for the production of four of such new artificial metalloenzymes, (A3_A3')Y26C and (A3_A3')F119C derived with Phen–Cu(II) and -Terpy–Cu(II), that were obtained by the covalent anchoring of either a copper(II)–phenanthroline cofactor or a copper(II)–terpyridine cofactor to either a F119C mutant or a Y26C mutant of the single chain A3_A3 protein. The strategy was

also successful to increase the yield of the reaction because with (A3_A3')F119C-Phen–Cu(II) as catalyst a $38 \pm 2\%$ yield could be obtained, but unfortunately in this case a very weak ee was found ($5 \pm 1\%$) as a result of both a bad control of the relative positioning of the two substrates in the wide pocket surrounding the Phen–Cu(II) cofactor and a too high flexibility of this cofactor. A better ee could be obtained with (A3_A3')F119C-Terpy–Cu(II) ($52 \pm 1\%$), but this time a low yield was found ($15 \pm 1\%$) resulting both from a lower flexibility of the bulky Terpy–Cu(II) cofactor and from a smaller sized pocket surrounding it, allowing a better control of the relative positioning of the two substrates. These results together with those reported earlier³ put in light that a good compromise between flexibility of the cofactor and size of the pocket surrounding it remain to be found to improve both the yield and the enantioselectivity of the reaction. Another important factor that has probably to be taken into account is the accessibility of the substrates to the catalytic pocket, as until now the A3 and A3_A3-derived biohybrids generally lead to much lower yields than some reported in the literature.^{7,16–18} Indeed, the conformational study and ligand binding study of A3_A3 shed light on the very tightly stabilized closed conformation in the bidomain single chain protein.⁵ Our results demonstrated that artificial metalloenzymes based on the A3_A3 protein modified with nitrogen donor ligands may be suitable for further catalyst optimization and constitute valuable tools toward more efficient and selective artificial biocatalysts (Table 3).

EXPERIMENTAL SECTION

Plasmid Construction. The genetic sequence coding for the protein α Rep A3 was issued from a combinatorial library of concatenated, repeated, and degenerated sequences. In order to mutate the first domain only, repetitions in nucleotide sequences were minimized to simplify site-directed mutagenesis. The sequence was therefore recoded using different codons for the same amino acid in consecutive repeats. The new nucleotide sequence was designed using the GenScript algorithm dedicated to codon optimization for *E. coli* expression as previously described.³ The bidomain genetic construction was recently described.⁵ Briefly, the synthetic gene of A3 was cloned by CPEC into pQE81 plasmid containing the original repeated sequence of A3 gene (named A3' in the mutated bidomain construction). The two genes were separated by a linker sequence coding for repeated serine and glycine residues (SGGGSGGGG). The gene also included a tobacco etch virus (TEV) protease cleavage site after the N-terminal hexahistidine tag. The final construct then encoded *HisTag-TEVsite-A3-(SG₄)₂-A3'*.

Directed Mutagenesis. Genes coding for the variants (A3_A3')F119C and (A3_A3')Y26C were obtained by site-directed mutagenesis using the QuikChange protocol with the following primers (F119C-FOR: 5'-CGGCGGCGTgcGCAT-TAGG; F119C-REV: 5'-CCTAATGcgcCGCCGCCG; Y41C-FOR: 5'-GCAGCGGCatgcGCGCTGGG; and Y41C-REV: 5'-CCCAGCGCgcaTGCCGCTG). The vector containing the A3_A3' gene was amplified by PCR with Phusion high-fidelity DNA polymerase (Thermo Scientific) as follows: initial denaturation at 98 °C for 1 min; 20 cycles including a denaturation step at 98 °C for 20 s, an annealing step at 65 °C for 15 s, and an extension step at 72 °C for 3 min. The final extension step was done at 72 °C for 10 min. The PCR product was treated with Dpn I and directly used for

transformation in *E. coli* XL1-Blue electrocompetent cells. For each mutagenesis, plasmids from five independent clones were purified using Nucleospin Plasmid extraction kit (Macherey Nagel). The sequences of the variants were confirmed by sequencing both DNA strands.

Protein Production. Chemocompetent *E. coli* M15 (Qiagen) bacteria were transformed with the desired plasmid and grown on 2YT-agar plates supplemented with ampicillin (100 $\mu\text{g}/\text{mL}$), kanamycin (50 $\mu\text{g}/\text{mL}$), and 1% glucose. One clone was picked and grown overnight at 37 $^{\circ}\text{C}$ in liquid 2YT medium supplemented with ampicillin (100 $\mu\text{g}/\text{mL}$) and kanamycin (50 $\mu\text{g}/\text{mL}$). The overnight preculture was used to inoculate 1 L of fresh medium distributed into two 2 L flasks. Cells were grown at 37 $^{\circ}\text{C}$ until $\text{OD}_{600\text{nm}}$ reached 0.8, protein expression was induced by the addition of 500 μM IPTG and cultures were further incubated at 37 $^{\circ}\text{C}$ for 4 h. Cells were harvested by centrifugation for 20 min at 5000g at 4 $^{\circ}\text{C}$ and resuspended into 40 mL of washing buffer (50 mM sodium phosphate pH 8; 300 mM NaCl; 20 mM imidazole). Suspended cells were stored at -80°C .

Protein Purification. Frozen cells from 1 L of bacterial culture were thawed and a tablet of EDTA-free anti-protease cocktail (Roche) was added. Cells were lysed by three 30 s cycles of sonication on ice and centrifuged 30 min at 11 000g at 4 $^{\circ}\text{C}$. His-tagged proteins were purified from the crude supernatant using nickel-affinity chromatography on a 4 mL Ni-NTA column (Macherey Nagel). The column was washed with 150 mL of washing buffer (50 mM sodium phosphate pH 8; 300 mM NaCl, 20 mM imidazole) and proteins were eluted in six 1 mL fractions in the elution buffer (50 mM sodium phosphate pH 8; 300 mM NaCl, 300 mM imidazole). A typical yield of about 100 mg of protein per liter of culture was obtained. Eluted fractions were analyzed by SDS-PAGE and pure fractions of proteins were pooled, diluted five times with 50 mM sodium phosphate buffer pH 8, 150 mM NaCl, and finally cleaved with His-tagged TEV protease obtained from the pRK793 vector (Addgene): purified His-tagged TEV protease and purified αRep bidomain protein with its cleavable His-Tag were mixed in a 1:100 ratio in a total volume of 50 mL and incubated overnight at 10 $^{\circ}\text{C}$ after the addition of 0.5 mM EDTA and 1 mM DTT. The mixture was then diluted twice with 50 mM sodium phosphate buffer pH 8, 150 mM NaCl leading to a final 30 mM imidazole concentration and loaded on another 4 mL Ni-NTA column equilibrated with the washing buffer. His-tagged TEV protease and uncleaved His-tagged αRep were retained on the column, while the flow through containing cleaved αRep protein was recovered, and concentrated resulting in a 5 mL solution that was loaded on a gel filtration column (HiLoad 16/600 Superdex 75, GE Healthcare) equilibrated with filtered coupling buffer (50 mM sodium phosphate pH 7.75, 150 mM NaCl). The coupling buffer was previously incubated for 1 h with Chelex 100 resin (Bio-Rad). Elution fractions after gel filtration were pooled and stored at 4 $^{\circ}\text{C}$ until chemical coupling. Purity was checked by SDS-PAGE and concentrations were obtained from the absorbance at 280 nm measured on a Nanodrop spectrophotometer (Thermo Scientific) using extinction coefficients at 280 nm calculated from the protein sequence (<https://web.expasy.org/protparam/>) (58 900 $\text{M}^{-1}\cdot\text{cm}^{-1}$ for A3_A3, 58 900 $\text{M}^{-1}\cdot\text{cm}^{-1}$ for (A3_A3')F119C and 57 410 $\text{M}^{-1}\cdot\text{cm}^{-1}$ for (A3_A3')Y26C). After the cleavage step, a typical yield of 40 mg of pure protein ready for the coupling reaction was obtained.

Monitoring of Thiols Accessibility. The accessibility of thiol groups of Cys119 and Cys26 in both variants was monitored using Ellman's test.²⁰ In a standard assay, 445 μL of reaction buffer (0.1 M sodium phosphate pH 8.0, 1 mM EDTA) was mixed with 5 μL of reagent solution [10 mM 5,5'-dithio-bis-(2-nitrobenzoic acid)] in a quartz cuvette. Absorbance at 412 nm was monitored using a JASCO spectrophotometer. Absorbance was recorded in the kinetic mode after the addition of 50 μL of a 100 μM protein solution until complete reaction. The final absorbance obtained at the plateau was noted A_{eq} and the initial absorbance measured before protein addition was noted A_0 . The reduced cysteine content was then calculated from the difference signal corrected by the dilution factor using an ϵ_{M} value of 14 150 $\text{M}^{-1}\cdot\text{cm}^{-1}$ at 412 nm.

Ligand Synthesis. Synthesis of 2-Bromo-N-(1,10-phenanthroline-5-yl)acetamide Hydrobromide (1). Compound (1) Was Synthesized According to Literature.⁷ 1,10-phenanthroline-5-amine (195 mg, 1 mmol) was solubilized in a 50 mL two-necked round-bottom flask filled with 30 mL of anhydrous CH_2Cl_2 and equipped with magnetic stirring and a condenser. Bromoacetyl bromide (104 μL , 1.2 mmol) was diluted in 10 mL anhydrous CH_2Cl_2 and added dropwise to the phenanthroline solution. The mixture was then refluxed overnight and the resulting suspension was filtered. Recrystallization was performed in MeOH, leading to the HBr salt of compound (1) as brownish crystals (yield 70%). ^1H NMR (300 MHz, $\text{DMSO}-d_6$): δ 10.86 (s, 1H), 9.32 (dd, $J = 4.5, 1.2$, 1H), 9.21 (dd, $J = 5, 1.2$, 1H), 9.09 (d, $J = 8.5, 1\text{H}$), 9 (d, $J = 8.5, 1\text{H}$), 8.54 (s, 1H), 8.20 (t, $J = 4.3, 1\text{H}$), 8.17 (t, $J = 4.3, 1\text{H}$), 4.34 (s, 2H) ppm.

Synthesis of 2-(2,2':6',2''-Terpyridine-4'-yloxy) Ethylamine (2). Compound (2) Was Synthesized According to Literature.²¹ 4'-Chloro-2,2':6',2''-terpyridine (159.7 mg) and ethanolamine (60 μL) were stirred in a suspension of powdered KOH (100 mg) in DMSO (2 mL) at 40 $^{\circ}\text{C}$ for 2 h. The reaction mixture was then added to 10 mL of CH_2Cl_2 and washed three times with water. The CH_2Cl_2 solution was dried over MgSO_4 and solvents evaporated. 2-(2,2':6',2''-Terpyridine-4'-yloxy) ethylamine (145 mg) was obtained as a light-yellow solid in 83% yield and used subsequently without further purification. ^1H NMR (360 MHz, CDCl_3): δ 3.17 (t, $J = 5.11$ Hz, 2H), 4.27 (t, $J = 5.11$ Hz, 2H), 7.33 (dd, $J_1 = 4.77$ Hz, $J_2 = 0.92$ Hz, 2H), 7.86 (td, $J_1 = 11.67$ Hz, $J_2 = 2$ Hz, 2H, $\text{H}_{4,4''}$), 8.03 (s, 2H), 8.61 (d, $J = 7.96$ Hz, 2H), 8.69 (d, $J = 3.85$ Hz, 2H) ppm. ESI-MS m/z : 293.13 (M), 315.12 (M + Na^+).

Synthesis of N-[2-([2,2':6',2''-Terpyridin]-4'-yloxy)ethyl]maleimide (3). 2-([2,2':6',2''-Terpyridin]-4'-yloxy)ethylamine (2) (136.4 mg) in 5 mL of acetone/water (2/1) was added dropwise to a solution of NaHCO_3 (118 mg) and *N*-methoxycarbonylmaleimide (127 mg) in 5 mL of acetone/water: 2/1 at 4 $^{\circ}\text{C}$, and the reaction mixture was stirred for 1 h 30 min. The solution was added to water (3 mL) and stirred for 2 h 30 min at room temperature. The obtained precipitate was collected and washed with water. Yield: 158.2 mg (80%); ^1H NMR (360 MHz, CDCl_3): δ 8.68 (d, $J = 4.83$ Hz, 2H), 8.60 (d, $J = 8.10$ Hz, 2H), 7.98 (s, 2H), 7.84 (td, $J = 7.72, 1.88$ Hz, 2H), 7.34 (dd, $J = 4.82, 1.22$ Hz, 2H), 6.75 (s, 2H), 4.40 (t, $J = 5.57$ Hz, 2H), 4.03 (t, $J = 5.55$ Hz, 2H) ppm. ESI-MS m/z : = 373.13 (M), 395.11 (M + Na^+).

Ligands Coupling. In both protein variants, the concentration of thiol groups obtained via Ellman's test was used as the cysteine concentration and thus of protein,

available for either phenanthroline or terpyridine coupling. Then, 1.2 equiv of compound 1, or 2 equiv of compound 3 were diluted in DMSO and mixed with 100 μM protein solution in coupling buffer (50 mM Na-phosphate pH 7.75, 150 mM NaCl) for a final DMSO concentration of 1% in the desired volume. The mixture was incubated at 4 $^{\circ}\text{C}$ overnight. The coupling reaction was then stopped by the addition of 0.5 mM EDTA and 1 mM DTT. The protein solution was concentrated to about 5 mL using a Vivaspin concentrator and loaded on a gel filtration column (HiLoad 16/600 Superdex 75, GE Healthcare) equilibrated in buffer A (20 mM MOPS 150 mM NaCl, pH7) previously incubated with Chelex 100 resin and filtered. The coupling efficiency was measured using Ellman's test and MALDI-ToF MS. The molar extinction coefficient at 280 nm was calculated by adding the calculated molar extinction coefficient of protein variants with the one measured for the phenanthroline or the terpyridine moiety. The latter was determined in a 70% ethanol solution of a known concentration of 1,10-phenanthroline and terpyridine, giving a molar extinction coefficient of 22 428¹ and 8000 $\text{M}^{-1}\cdot\text{cm}^{-1}$, respectively.

Mass Spectrometry Analysis. Protein solutions were desalted on a Zeba Spin column (Thermo) against 10 mM NH_4HCO_3 buffer pH 8 and then diluted in a sinapinic acid matrix for MALDI-ToF MS analysis in linear mode by using a MALDI-ToF/ToF UltrafleXtreme spectrometer (Bruker Daltonics).

Metalation of the apo-Biohybrids. Aqueous $\text{Cu}(\text{NO}_3)_2\cdot 3\text{H}_2\text{O}$ (1 μL , 10 mM) was added to 100 μL of 100 μM solution of either A3_A3, (A3_A3')F119C-Phen, (A3_A3')-F119C-Terpy, (A3_A3')Y26C-Phen or (A3_A3')Y26C-Terpy in buffer A and gently yet thoroughly mixed by pipetting in order to obtain a 1:1 copper(II)/protein ratio. The solutions were kept on ice in a cold room overnight.

Analytical Size Exclusion Chromatography. Analytical size exclusion chromatography was done essentially as described⁷ with an ÄKTA Purifier (GE Healthcare) system on a Superdex 75 10/300 GL column (flow rate 1 $\text{mL}\cdot\text{min}^{-1}$) equilibrated with buffer A. Solutions containing sample proteins were injected and elution profiles were recorded by following the absorbance at 280 nm.

Molecular Modeling. The geometry of the Trojan horse models (cofactor, substrate bound cofactor, and product models) were optimized with Gaussian 09 (Gaussian 09, revision D.01. Gaussian, Inc.: Wallingford, CT, 2010) using the wave function B3LYP-D3²² and the SMD continuum model for water²³ (see Supporting Information for further details).

The dimeric bio-assembly of the X-ray structure of the A3_A3 bidomain protein (PDB ID 6FT5) was used to model the different biohybrid variants (A3_A3')F119C and (A3_A3')Y26C by mutating the relevant side chains. The positions of the bidomain A3_A3' protein were numbered starting from residue number 3, as the first two residues of the sequence were lacking in the crystal structure. The numbers of the mutated positions were thus in accordance with those described previously for the A3 homodimer.³ All the structural handling performed at the beginning of the study was carried out by means of UCSF Chimera embedded software.²⁴

Covalent protein–ligand docking assays were performed with the program GOLD (version 5.2)²⁵ using the GoldScore scoring function²⁶ and considering the bond between the cofactor and the sulfur of the mutated cysteine as the covalent anchoring. An evaluation sphere of 20 Å radius centered on the

sulfur atom of the mutated cysteines was used for each simulation. In all the calculations, the Trojan horses and the linker were treated as flexible structures.

The solutions were analyzed by means of GaudiView GUI extension for Chimera²⁷ taking into account three main criteria, the maximum value of the GoldScore *Fitness* (F_{max}), the position in the global energy ranking, and the population of the solution clusters.

General Procedure for the Catalysis of Diels–Alder Cyclization Reactions. All solutions were kept on ice and reaction runs were performed in duplicate in 2 mL Eppendorf conical tubes. In these tubes, the following were added successively: 150 μL of the 100 μM solution of copper(II)-biohybrid prepared previously (final concentration 30 μM), 5 μL of a stock solution of 100 mM azachalcone 4 (final concentration 1 mM, 33 equiv), 5 μL of a 187 mM diphenylsulfone solution used as an internal standard (final concentration 1.87 mM) in acetonitrile, and finally 1.4 μL of pure freshly distilled cyclopentadiene 5 (final concentration 34 mM, 1133 equiv) into buffer A for a final volume of 500 μL . The tubes were then placed in the ThermoMixer comfort and stirred at 500 rpm for 48 h at 4 $^{\circ}\text{C}$. Ethyl acetate (3 \times 500 μL) was used for extracting the reaction mixture. After centrifugation, organic layers were collected and dried over MgSO_4 . HPLC analyses were then performed at 20 $^{\circ}\text{C}$ on an Agilent Technologies 1260 Infinity apparatus equipped with a Lux Cellulose-1 (250 \times 4.6 mm, 5 μm) and *n*-hexane/isopropanol (99/1) as the mobile phase at a 0.5 $\text{mL}\cdot\text{min}^{-1}$ flow rate (DAD: 270 nm for 6 and diphenylsulfone, 300 nm for 4). The peaks corresponding to the four stereoisomers were identified by comparison with those of authentic samples: retention times: 15.4 and 17.0 min (6 exo isomers), 18.7 and 24.6 min (6 endo isomers), 29.1 min (starting material, 4), and 61 min (internal standard, diphenylsulfone). The stereoisomers were assessed after circular dichroism measurements.⁷

■ ASSOCIATED CONTENT

📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsomega.8b03448.

Molecular modeling; GoldScore values for different docking simulations; and Cartesian coordinates of the DFT optimized structures (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: jean-pierre.mahy@u-psud.fr (J.-P.M.).

*E-mail: agathe.urvoas@u-psud.fr (A.U.).

*E-mail: remy.ricoux@u-psud.fr (R.R.).

ORCID

Giuseppe Sciortino: 0000-0001-9657-1788

Rémy Ricoux: 0000-0002-3464-7895

Author Contributions

||T.D.M. and K.K. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Schwizer, F.; Okamoto, Y.; Heinisch, T.; Gu, Y.; Pellizzoni, M. M.; Lebrun, V.; Reuter, R.; Köhler, V.; Lewis, J. C.; Ward, T. R.

Artificial Metalloenzymes: Reaction Scope and Optimization Strategies. *Chem. Rev.* **2018**, *118*, 142–231.

(2) Urvoas, A.; Guellouz, A.; Valerio-Lepiniec, M.; Graille, M.; Durand, D.; Desravines, D. C.; van Tilbeurgh, H.; Desmadril, M.; Minard, P. Design, Production and Molecular Structure of a New Family of Artificial Alpha-helicoideal Repeat Proteins (α Rep) Based on Thermostable HEAT-like Repeats. *J. Mol. Biol.* **2010**, *404*, 307–327.

(3) Di Meo, T.; Ghattas, W.; Herrero, C.; Velours, C.; Minard, P.; Mahy, J.-P.; Ricoux, R.; Urvoas, A. ARep A3: A Versatile Artificial Scaffold for Metalloenzyme Design. *Chem.—Eur. J.* **2017**, *23*, 10156–10166.

(4) Guellouz, A.; Valerio-Lepiniec, M.; Urvoas, A.; Chevrel, A.; Graille, M.; Fourati-Kammoun, Z.; Desmadril, M.; van Tilbeurgh, H.; Minard, P. Selection of Specific Protein Binders for Pre-Defined Targets from an Optimized Library of Artificial Helicoideal Repeat Proteins (AlphaRep). *PLoS One* **2013**, *8*, No. e71512.

(5) Léger, C.; Di Meo, T.; Aumont-Nicaise, M.; Velours, C.; Durand, D.; Li de la Sierra-Gallay, I.; van Tilbeurgh, H.; Hildebrandt, N.; Desmadril, M.; Urvoas, A.; Valerio-Lepiniec, M.; Minard, P. Ligand-Induced Conformational Switch in an Artificial Bidomain Protein Scaffold. *Sci. Rep.* **2019**, *9*, 1178.

(6) Himiyama, T.; Sauer, D. F.; Onoda, A.; Spaniol, T. P.; Okuda, J.; Hayashi, T. Construction of a Hybrid Biocatalyst Containing a Covalently-Linked Terpyridine Metal Complex within a Cavity of Aponitrobindin. *J. Inorg. Biochem.* **2016**, *158*, 55–61.

(7) Bos, J.; Fusetti, F.; Driessen, A. J. M.; Roelfes, G. Enantioselective Artificial Metalloenzymes by Creation of a Novel Active Site at the Protein Dimer Interface. *Angew. Chem., Int. Ed.* **2012**, *51*, 7472–7475.

(8) Ellman, G. L. Tissue Sulfhydryl Groups. *Arch. Biochem. Biophys.* **1959**, *82*, 70–77.

(9) Funel, J.-A.; Abele, S. Industrial Applications of the Diels-Alder Reaction. *Angew. Chem., Int. Ed.* **2013**, *52*, 3822–3863.

(10) Gregoritz, M.; Brandl, F. P. The Diels-Alder reaction: A powerful tool for the design of drug delivery systems and biomaterials. *Eur. J. Pharm. Biopharm.* **2015**, *97*, 438–453.

(11) Klas, K.; Tsukamoto, S.; Sherman, D. H.; Williams, R. M. Natural Diels-Alders: Elusive and Irresistible. *J. Org. Chem.* **2015**, *80*, 11672–11685.

(12) Nicolaou, K. C.; Hale, C. R. H.; Nilewski, C.; Ioannidou, H. A. Constructing Molecular Complexity and Diversity: Total Synthesis of Natural Products of Biological and Medicinal Importance. *Chem. Soc. Rev.* **2012**, *41*, 5185.

(13) Filice, M.; Romero, O.; Gutiérrez-Fernández, J.; de las Rivas, B.; Hermoso, J. A.; Palomo, J. M. Synthesis of a Heterogeneous Artificial Metalloenzyme with Chimeric Catalytic Activity. *Chem. Commun.* **2015**, *51*, 9324–9327.

(14) Ghattas, W.; Dubosclard, V.; Wick, A.; Bendelac, A.; Guillot, R.; Ricoux, R.; Mahy, J.-P. Receptor-Based Artificial Metalloenzymes on Living Human Cells. *J. Am. Chem. Soc.* **2018**, *140*, 8756–8762.

(15) Himiyama, T.; Taniguchi, N.; Kato, S.; Onoda, A.; Hayashi, T. A Pyrene-Linked Cavity within a β -Barrel Protein Promotes an Asymmetric Diels-Alder Reaction. *Angew. Chem., Int. Ed.* **2017**, *56*, 13618–13622.

(16) Ghattas, W.; Cotchico-Alonso, L.; Maréchal, J.-D.; Urvoas, A.; Rousseau, M.; Mahy, J.-P.; Ricoux, R. Artificial Metalloenzymes with the Neocarzinostatin Scaffold: Toward a Biocatalyst for the Diels-Alder Reaction. *ChemBioChem* **2016**, *17*, 433–440.

(17) Filice, M.; Romero, O.; Gutiérrez-Fernández, J.; de las Rivas, B.; Hermoso, J. A.; Palomo, J. M. Synthesis of a Heterogeneous Artificial Metalloenzyme with Chimeric Catalytic Activity. *Chem. Commun.* **2015**, *51*, 9324–9327.

(18) Reetz, M. T.; Jiao, N. Copper-Phthalocyanine Conjugates of Serum Albumins as Enantioselective Catalysts in Diels-Alder Reactions. *Angew. Chem., Int. Ed.* **2006**, *45*, 2416–2419.

(19) Osseili, H.; Sauer, D. F.; Beckerle, K.; Arlt, M.; Himiyama, T.; Polen, T.; Onoda, A.; Schwaneberg, U.; Hayashi, T.; Okuda, J. Artificial Diels-Alderase based on the transmembrane protein FhuA. *Beilstein J. Org. Chem.* **2016**, *12*, 1314–1321.

(20) Eyer, P.; Worek, F.; Kiderlen, D.; Sinko, G.; Stuglin, A.; Simeon-Rudolf, V.; Reiner, E. Molar Absorption Coefficients for the Reduced Ellman Reagent: Reassessment. *Anal. Biochem.* **2003**, *312*, 224–227.

(21) Maayan, G.; Yoo, B.; Kirshenbaum, K. Heterocyclic Amines for the Construction of Peptoid Oligomers Bearing Multi-Dentate Ligands. *Tetrahedron Lett.* **2008**, *49*, 335–338.

(22) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate *Ab Initio* Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(23) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.

(24) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera?A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

(25) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking 1 Edited by F. E. Cohen. *J. Mol. Biol.* **1997**, *267*, 727–748.

(26) Jones, G.; Willett, P.; Glen, R. C. Molecular Recognition of Receptor Sites Using a Genetic Algorithm with a Description of Desolvation. *J. Mol. Biol.* **1995**, *245*, 43–53.

(27) Rodríguez-Guerra Pedregal, J.; Sciortino, G.; Guasp, J.; Muncioy, M.; Maréchal, J.-D. GaudiMM: A Modular Multi-Objective Platform for Molecular Modeling. *J. Comput. Chem.* **2017**, *38*, 2118–2126.