FEBS Letters

FEBS PRESS science publishing by scientists

Engineering protein fragments via evolutionary and protein–protein interaction algorithms: *de novo* design of peptide inhibitors for F_OF_1-ATP synthase

Yasser B. Ruiz-Blanco[1,*], Luis Pablo Ávila-Barrientos[1], Enrique Hernández-García[1], Agostinho Antunes[2,3], Guillermin Agüero-Chapin[2,3] and Enrique García-Hernández[1]

1 Instituto de Química, Universidad Nacional Autónoma de México, México
2 CIMAR/CIIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Terminal de Cruzeiros do Porto de Leixões, Portugal
3 Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Portugal

Correspondence
G. Agüero-Chapin, CIMAR/CIIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Terminal de Cruzeiros do Porto de Leixões, Av. General Norton de Matos, s/n, Porto 4450-208, Portugal
Tel: (+351) 22 340 18 00
E-mail: gchapin@ciimar.up.pt
and
E. García-Hernández, Instituto de Química, Universidad Nacional Autónoma de México, Ciudad Universitaria, Ciudad de México 04510, México
Tel: +52 (55) 56 22 22 24
E-mail: egarciah@unam.mx

Present address
*Computational Biochemistry, Center of Medical Biotechnology, University of Duisburg-Essen, Essen, 45127, Germany

Yasser B. Ruiz-Blanco and Luis Pablo Ávila-Barrientos contributed equally to this article

(Received 24 September 2020, revised 23 October 2020, accepted 30 October 2020, available online 19 November 2020)

doi:10.1002/1873-3468.13988

Edited by Qinghua Cui

Enzyme subunit interfaces have remarkable potential in drug design as both target and scaffold for their own inhibitors. We show an evolution-driven strategy for the *de novo* design of peptide inhibitors targeting interfaces of the *Escherichia coli* FoF1-ATP synthase as a case study. The evolutionary algorithm ROSE was applied to generate diversity-oriented peptide libraries by engineering peptide fragments from ATP synthase interfaces. The resulting peptides were scored with PPI-Detect, a sequence-based predictor of protein–protein interactions. Two selected peptides were confirmed by *in vitro* inhibition and binding tests. The proposed methodology can be widely applied to design peptides targeting relevant interfaces of enzymatic complexes.

Keywords: peptide library; PPI-Detect; protein interfaces; ROSE; sequence evolution

The last couple of decades have witnessed a fast increase in the interest of peptides as therapeutic tools, whether as active ingredients or as drug–peptide complexes [1]. While previously considered troublesome drug candidates, today there are over 60 approved peptide drugs, and there are hundreds in

Abbreviations
AMPs, antimicrobial peptides; DEL, DNA-encoded library; MST, microscale thermophoresis; PPI, protein–protein/peptide interaction; ROSE, (Random Model of Sequence Evolution); SME, simulated molecular evolution.

clinical or preclinical trials [2]. Peptide drugs typically have low toxicity and are capable of extremely tight binding to their targets, and cell permeable peptides have been used as carriers for intracellular acting molecules [2]. Active areas of drug development for peptides include oncology [3], metabolic [4], vascular [5], and, with very promising results, microbial diseases [6]. Currently, most available peptide drugs are natural or derived thereof [1].

Commonly, the exploration of the structural space of peptide candidates is performed by the generation of combinatorial peptide libraries either by chemical synthesis or by biological intervention [7]. For the chemical synthesis, the 'split-and-mix' approach is the most popular, either by using traditional solid-phase peptide synthesis [8] or by the more recent DNA-encoded library (DEL) platform [9]. In addition, biological methods are also frequently used to produce peptide libraries by directed evolution approaches, either by subjecting a gene of interest to rounds of site-directed mutagenesis or by random mutations in order to release genetic variants that generate peptide diversification [7].

Despite the advances in methods for generating peptide libraries as well as for *in vitro* high-throughput screening analyses [7], they are still notoriously time-consuming and expensive. Commonly, *in silico* approaches are designed to produce random peptide libraries by performing stochastic substitutions in the different positions of an amino acid sequence [10,11]. Other algorithms have been gradually incorporated to guide the computational generation of peptides carrying certain secondary structures, such as amphipathic helices, kinked amphipathic helices, structures aimed to interact with lipid membranes, and others retaining symmetric features [12]. However, evolutionary algorithms have been scarcely applied to either generate peptides libraries from putative hits or in hit-to-lead generation processes [13].

Here, we propose the design, guided by evolutionary rules, of peptide libraries derived from short fragments of protein regions relevant for the function or modulation of the protein. The aim of this approach was to generate diversity-oriented libraries, where the exploration of the structural space is controlled by evolutionary parameters such as tree topology, evolutionary distance, mutation rate, insertions, and deletions, among others. The resulting peptide libraries can be analyzed with similarity-based search engines (alignment-based methods) or with alignment-free models in order to identify new candidates with the targeted biological activity. For the evolution-generated peptide libraries, there are several algorithms with potentialities to be applied to this aim. They were originally developed to evaluate the accuracy of multiple sequence alignment and phylogenetic reconstruction tools by generating sets of related simulated protein sequences from known phylogenies. Some examples of evolutionary algorithms are as follows: ROSE (Random Model of Sequence Evolution) [14], SIMPROT (Simulation Protein Evolution) [15], and INDELible (Insertions and Deletions Simulator) [16]. Here, we selected ROSE to generate diversity-oriented libraries from root peptides, because it offers a simple and versatile evolutionary algorithm that allows users to set suitable parameter values for the generation of thousands of related peptides at no demanding computer resources (Fig. S1). This new application of evolutionary algorithms for sequence simulations is a fundamental change in the rational strategies to explore the structural space in the generation of synthetic peptide libraries.

The virtual screening of large peptide libraries can benefit from the use of sequence-based protein–protein/peptide interaction (PPI) predictors. Given the amino acid sequences are the sole inputs of these methods, they leverage a large applicability together with fast execution rates [17]. The main approaches used to develop sequence-based PPI predictors encompass motif/domain-based searches, global sequence similarity, and codon usage in genes encoding interacting and noninteracting proteins [18]. Comprehensive revisions on this topic, as well as a list of available predictors of PPI sites, have been published by Ding and Kihara [18] and Casadio and colleagues [19]. The main shortcoming faced by sequence-based PPI predictors has been their poor precision, originated by an overproduction of false-positive hits [20]. Recently, Romero-Molina *et al.* [17] introduced PPI-Detect, a machine-learning-based predictor of PPI using domain information. PPI-Detect rendered a significantly increased precision in different test sets, which obeys to the incorporation of a reliable set of noninteracting domains in its training data, as well as a powerful numerical algorithm to encode and extract diverse features from amino acid sequences. In addition, PPI-Detect was challenged with the discrimination of active from inactive derivatives of EPIX4, an endogenous peptide inhibitor of the GPCR CXCR4 [17]. There, this program also showed a high success rate. Such results predicting interactions at fragment/peptide level make PPI-Detect a particularly suitable sequence-based predictor of PPI for the massive screening peptide binders of a certain target.

Among the molecules susceptible of being modulated therapeutically by peptides, F_OF_1-ATP synthase,

also called ATP synthase, has risen to prominence in recent years [21–23]. It is composed of a membrane sector, F_O, that transduces the proton motive force into rotation, and a soluble sector, F_1, that contains its three catalytic sites. In *Escherichia coli* and other bacteria, it is composed of eight types of subunits with F_O:$a_1b_2c_{10\text{-}12}$ and F_1:$\alpha_3\beta_3\gamma\delta\epsilon$ stoichiometries [24,25]. During rotation, numerous inter- and intracatenary interactions are coordinately formed and disrupted, property that highlights the importance of the interfaces in the communication and cooperativity among ATP synthase's subunits [26]. Additional to its most common locations (the inner membranes of mitochondria, bacteria, and chloroplast thylakoids), ATP synthase is found on the plasma membrane of several mammalian cell types (hepatocytes, adipocytes, and endothelial cells), playing crucial roles in regulating many cellular processes in humans [27]. These roles, as well as its involvement in various genetic diseases [28], have made human ATP synthase an attractive target for therapeutic molecules [29]. In addition, ATP synthase has been proven as a promising target for the development of new antibiotics [30]. This breakthrough has led to an accelerated search for compounds that could be used as new ATP synthase-targeting antibiotics [22]. The ongoing search has also included many natural peptides that show appealing pharmaceutical activities [29]. Despite the variety of exogenous peptide inhibitors that evidence the high 'druggability' of this enzyme, which makes it an attractive target to develop new antimicrobial peptides (AMPs), to the best of our knowledge no inhibitory peptide derived from ATP synthase fragments and targeting the same enzyme have been previously designed [22].

Here, we provide a new *in silico* strategy guided by evolutionary and machine learning-based methods for the *de novo* design of peptide inhibitors by engineering protein fragments of F_OF_1-ATP synthase. This strategy permits exploring a relevant structural space around root peptides with potential protein binding properties. The role of interactions at the interfaces of protein subunits in the structural flexibility of complex enzymatic systems, as well as in the stabilization of the inhibitors binding, has been demonstrated [31,32]. In this study, we selected fragments (peptides), ranging between 20 and 40 residues, which belong to interfaces between ATP synthase subunits. Given the involvement of these peptides in protein–protein interactions relevant for the functioning of the enzyme, we hypothesize that is possible to extract and tailor *in silico* such fragments, and turn them into effective modulators of those original protein–protein interactions. Such a concept, aimed at functionalizing a fragment of the enzyme as a novel inhibitor, constitutes a *de novo* design of bioactive peptides. To address this goal, we further envisioned the application of evolutionary algorithms together with an accurate machine learning-based predictor of protein–protein interactions, to rationally generate and screen a diversity-oriented peptide library (Fig. 1).

Materials and methods

Dataset

Fragments of interfaces involved in protein–protein interactions between subunits of ATP synthase were extracted subdivided into four classes: Class-1 are interfaces from the catalytic domain of the ATP synthase, taken from the structure with PDB ID 1E79, chains ID: ABCDEF (nine peptide fragments). Class-2 groups peptides from the rotor of the enzyme, extracted from the structure with PDB ID 1E79, chains ID: GHI (six peptide fragments). Class-3 gathers fragments from the hydrophilic portion of the stalk, taken from the PDB ID 2CLY (seven peptide fragments). Class-4 corresponds to fragments from the transmembrane domain, extracted from the PDB ID 1C17 (seven peptide fragments). All the extracted fragments possess sequence lengths between 20 and 90 amino acids (File S1).

Multiple sequence alignments to build consensus root sequences

Each peptide class was aligned independently using MAFFT (Multiple Alignment using Fast Fourier Transform) v7.215, a high-speed multiple sequence alignment. The iterative refinement method FFT-NS-i was used with default parameters [33]. Multiple sequence alignments (MSA) were used to extract consensus sequences by estimating the frequency of each residue at every column of the alignment. The residues above an identity threshold of 60% conformed the positions in the consensus sequence. The amino acid sequences of all the fragments extracted from the ATP synthase as well as their corresponding MSA are summarized in File S1. The obtained consensus sequences for each class were subsequently used as root peptides to feed ROSE, to generate the final peptide library.

ROSE: Random model of sequence evolution

The ROSE program (https://bibiserv.cebitec.uni-bielefeld.de/rose) creates a family of related peptides from a root peptide by applying a probabilistic model of evolution considering events such as insertion, deletion, and substitution of characters guided by the topology and branch lengths of

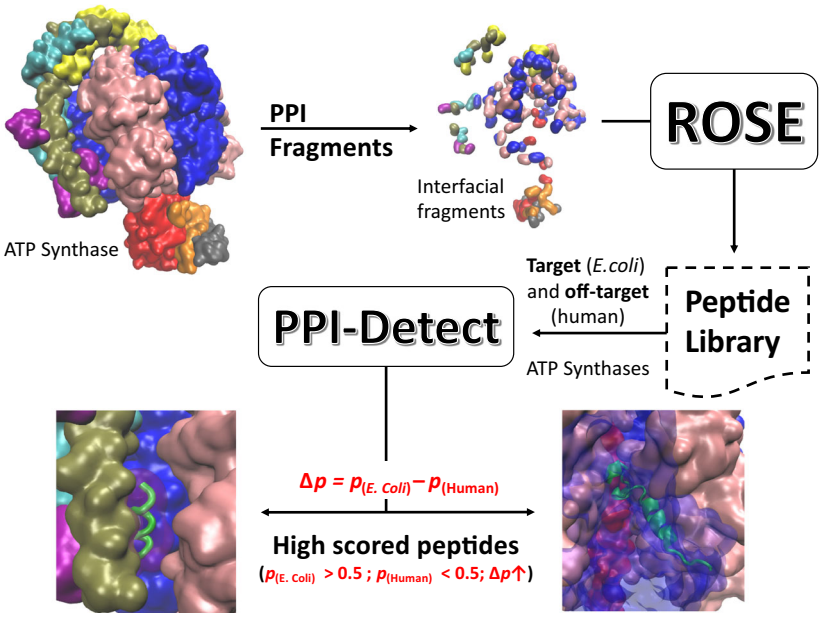

Fig. 1. Schematic representation of the applied *de novo* design strategy. The procedure starts by extracting multiple fragments involved in protein–protein interactions (PPI) from ATP synthase structures. The ROSE program is used for generating an evolution-driven peptide library from native fragments of ATP synthases. The program PPI-Detect is employed for the virtual screening of this library to identify potential binders of the target ATP synthase.

a predefined evolutionary tree. ROSE was calibrated, so that the generated peptides retained at least 70% of identity with the corresponding root sequence. ROSE's internal parameters were tuned as follows: the binary mutation guide trees with 1023 nodes and depth $k = 9$, with average distance (d_{av}) = 5–20 PAMs (Fig. S1). In addition to d_{av}, the diversity of the resulting peptides also depends on the root sequence, which is represented by a mutation probability vector where each position/residue is weighted by variability or conservation degree shown in the sequence consensus ['zero' value indicates no mutations (high conservation degree) while 'one' value represents high mutation probability]. An example of an input file for ROSE to generate a set of related peptides is shown in File S2.

PPI-Detect: sequence-based predictor of protein–protein interactions

PPI-Detect is a user-friendly web application [17], freely available within the PROTDCAL-SUITE [34] (https://protdcal.zmb.uni-due.de/), which allows predicting whether two proteins, or protein–peptide pairs, will interact or not, using exclusively their primary structure information. PPI-Detect integrates a novel numeric encoding method of pairs of peptides sequences from ProtDCal [35] and a Support Vector Machine (SVM) classifier based on information contained in two databases of protein domain–domain interactions 3Did [36] and i-Pfam [37], as well as in Negatome 2.0 [38], a curated database of domains with very low chances of forming stable complexes. In this context, PPI-Detect was used to screen the libraries regarding the peptide interactions with both domains of the *E. coli* and human ATP synthases.

In vitro inhibition test of the ATPase activity of *E. coli* F_1 sector (*Ec*F_1)

Unless stated otherwise, all the chemicals were from Sigma Chemical (St. Louis, MO, USA). Recombinant wild-type *Ec*F_1 was expressed in cells of *E. coli* strain DK8 transformed with the pBWU13.4 plasmid and purified as described elsewhere [39,40]. Protein concentrations were determined using the Pierce BCA Protein Assay Kit (Thermo-Fisher, Rockford, Illinois, USA). ATPase activities were determined as reported elsewhere [41], following inorganic phosphate concentration using the malachite green phosphate assay [42] in 96-well microplates. All peptides were previously incubated with 1.1 μM *Ec*F_1 for 1 h at 30 °C, in the presence of 1 mM ATP, 1 mM Mg(II), 50 mM TRIS-SO_4, pH 8.0. The concentration of inhibitor required to achieve a 50% reduction in enzymatic activity, IC_{50}, was obtained using the Hill equation:

$$\frac{v_i}{v_0} = \frac{IC_{50}^h}{IC_{50}^h + [\text{Peptide}]^h} + V_r,$$

where v_0 and v_i are the initial catalytic velocities in absence and in presence of a given concentration of peptide, [Peptide]. h is the Hill coefficient and V_r is the residual velocity at a maximal inhibition of the peptide.

Microscale thermophoresis

Microscale thermophoresis (MST) was used to measure the interaction between *Ec*F1 and the designed peptides in a Monolith NT.115 instruments (NanoTemper Technologies GmbH) [43]. The enzyme was labeled with the fluorescent

red dye NT-647. A peptide stock was serially diluted (1 : 1) with $Ec\mathrm{F}_1$ at a constant concentration of 50 nM in 50 mM TRIS-SO_4, pH 8.0M. In all experiments, after a short incubation, the samples were loaded into Monolith™ NT.115 premium Treated Capillaries. Thermophoresis was defined as the ratio of signals within a 1 s time window ~ 1 s and ~ 20 s after the laser was switched on. The change in the normalized fluorescence (F_{norm}), which is defined as the ratio of the fluorescence values prior to and after IR laser activation, was fitted to

$$F_{\mathrm{norm}} = U + \frac{B-U}{1+\left(\frac{\mathrm{EC}_{50}}{[\mathrm{Peptide}]}\right)^{h}},$$

where EC_{50} is the peptide concentration that gives half-maximal response, and U and B represent signals of the enzyme unbound and bound states, respectively.

Results

De novo design strategy and results outlined

The approach presented in this study was implemented as follows. The fragments from the ATP synthases were mutated by ROSE to build an extended peptide library as a source of new inhibitors targeting the interfaces of the above-mentioned subunits (Fig. S1). The resulting libraries were then screened to identify putative binders of *E. coli* ATP synthase, by using the program PPI-Detect (https://ppi-detect.zmb.uni-due.de/), a support vector machine (SVM)-based model trained to identify whether two protein domains will interact, and also assessed in the prediction of protein–peptide interactions [17]. In order to reduce the likelihood of cross-species reactivity, the interaction probability of peptide candidates with domains of human ATP synthase was also scored. Finally, two peptides selected among the top-ranked candidates, that is those showing high score with *E. coli* and low with human ATP synthase, were synthetized and analyzed with *in vitro* inhibition and binding assays. Notably, both peptides, labeled as P1 and P2, showed concentration-dependent inhibition of the *E. coli* ATP synthase in the micromolar order. P1 and P2 were derived from the hydrophilic portion of the stalk chain and from the catalytic domains of the ATP synthase, respectively (Fig. 2).

Generation of the peptide library by using an evolutionary algorithm

Multiple sequence alignments were carried out on each of the four sets of fragments extracted from interfaces between subunits of the ATP synthase (File S2). The resulting consensus peptides, with sequences between 14 and 41 residues, are summarized in Table S1.

The probabilistic model of evolution implemented in ROSE was applied to generate sets of mutants from each of the consensus sequences summarized in Table S1. ROSE parameters were calibrated so that each set of mutants shared sequence identities equal or higher than 70%. These resulting sets were fused in an initial library of 5428 peptides. Subsequently, the redundancy within this starting library was eliminated by using CD-Hit [44] at a maximum identity cut off of 80%, which resulted in 360 final candidates in the peptide library. In this way, the evolutionary algorithm used for generating new peptides provided rationality to the exploration of the structural space of peptides during the identification process of hits in drug design.

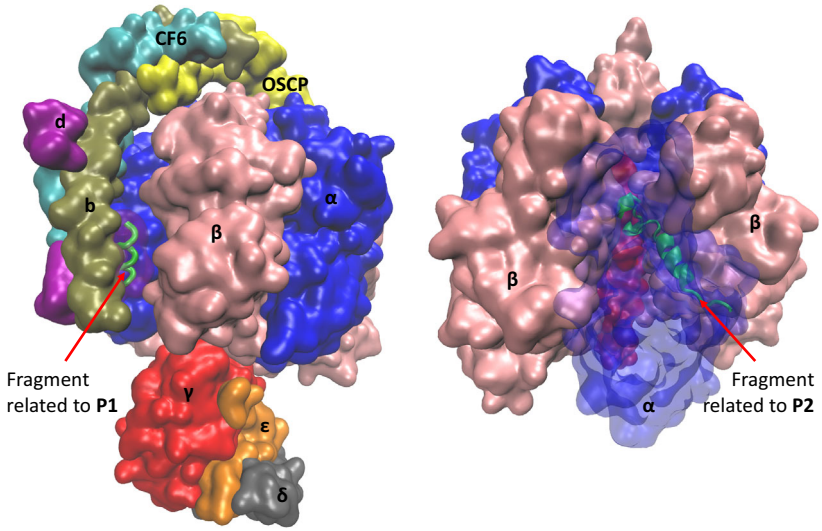

Fig. 2. Model representation of the native fragments from ATP synthases which are related to the identified inhibitors P1 and P2. The fragments are shown in green cartoons. Schematic representations of the ATP synthase structure are presented in surface colored by subunits: α, pink; β, blue; γ, red; δ, gray; ε, orange; subunit OSCP, yellow; *b* subunit, tan; *d* subunit, purple; and coupling factor 6, cyan. The chains containing the highlighted fragments are depicted with transparent surfaces. The structure on the left corresponds to the PDB ID 2WSS [67] and the heptamer on the right to the PDB ID 1NBM [68].

Virtual screening of the peptide libraries using PPI-Detect

PPI-Detect [17] was used to score the interaction likelihood between each peptide and all the main subunits in *E. coli*'s (*Ec*) and in the human's (*Hs*) F_1Fo-ATP synthase. Finally, eight candidates were extracted according to three selection criteria: (a) Peptides with maximum interaction likelihood with *Hs*F_1Fo below 0.5. (b) Peptides with maximum interaction likelihood with *Ec*F_1Fo above 0.5. (c) The difference between the interaction probabilities with *E. coli* and human enzymes is at least 0.1 (Table 1).

Three peptides, P1, P2, and P6, out of the eight final candidates obtained from the virtual screening show a net positive charge, which is a common feature among antibacterial and antitumoral peptides. Among these three candidates, P1 and P2 have the most negative GRAVY indices, which suggest that they have better solubility than P6. In addition, they also show the highest predicted value of helical content, what is related to the structure of the parent fragment in the ATP synthase (Fig. 2). We therefore selected these two promising peptides for synthesis and assessment of their inhibitory power against *Ec*F_1. Since such peptides did not show significant similarities with any member of the most comprehensive AMP database reported up to date [45], they can be considered as '*de novo*-designed' inhibitors.

In vitro inhibition of *Ec*F_1 by the designed peptides

Peptides P1 and P2 were synthetized and their inhibition activities were tested against *Ec*F_1 by following inorganic phosphate liberation. Figure 3A shows the inhibitory effect determined at a single peptide concentration of 50 μM. Clearly, the two peptides inhibited the enzyme far beyond experimental uncertainty. To further characterize the inhibitory activities, dose–response measurements were carried out (Fig. 3B). Nonlinear analysis of these data using the Hill equation showed inhibitory potencies in the micromolar range for both peptides ($IC_{50} = 46 \pm 2$ and 53 ± 2 μM for P1 and P2, respectively). The Hill number obtained for P1 ($h = 2.91 \pm 0.36$) revealed a positive cooperativity in the interaction of this peptide with the enzyme, while that obtained for P2 ($h = 0.97 \pm 0.2$) indicated an apparent 1:1 stoichiometry. Interestingly, the residual velocity (V_r) was nearly zero for both peptides, showing that they act as dead-end inhibitors.

In addition, we used MST as an orthogonal method to verify the interaction of *Ec*F1 and the two peptides (Fig. 4). As an example of MST results, Fig. 4A shows the traces obtained for P1. Titration of the nonfluorescent ligand resulted in a gradual change in MST signal. The dose–response curves yielded results in good agreement with those obtained by enzyme kinetic assays (Fig. 4B,C). P1 yielded EC_{50} and h values of 12.5 ± 0.2 μM and 1.6 ± 0.1, respectively, again showing a positive cooperativity in the interaction. In the case of P2, $EC_{50} = 90 \pm 30$ μM, while h was set to 1 due to the lack of convergence in the data fitting process. In summary, the experimental results gave clear support for a direct inhibitory interaction of the two designed peptides against *Ec*F_1.

Discussion

F_OF_1-ATP synthase

F_OF_1-ATP synthase has a sophisticated rotary mechanism that involves the formation and disruption of

Table 1. Summary of the final peptides, and their interaction scores, from the virtual screening of the evolutionary library.

Name	Sequence	Class	Score[a]	Charge[b]	GRAVY[c]	Helicity[d] (%)
P1	**SLKEIQEAIDLDRELPKLMQKPPAIDWKYGKANMAKAPNV**	**Class-3**	**0.12**	**1**	**−0.675**	**60.0**
P2	**ARQVSALHQRIPSAVGYQPTISFAHFGKDLDALTKEERLT**	**Class-1**	**0.1**	**1**	**−0.370**	**55.0**
P3	TLEKPKFEALKVPIPEDLDYKYTAQVDAEPKE	Class-3	0.17	−3	−0.903	0.0
P4	AREVSALLQRIPSAVGYQPTIIFGHFGADLDATTEDERLT	Class-1	0.14	−3	−0.077	47.5
P5	AREVSPLLQREPSAVGYQPTISFGGFGSDLDALTENERLT	Class-1	0.13	−3	−0.378	40.0
P6	SLKEIQEAIDLARELAKLKQKPPVIDWMYGKATMAQAPSV	Class-3	0.11	1	−0.250	55.0
P7	LVIFNRDQQRPTRKSDYSVIAAELLADEQEKATL	Class-2	0.1	−1	−0.576	47.1
P8	LVTFNRDEQRPNHKADYSDIAAQLLRDEQEKGYL	Class-2	0.1	−2	−1.191	44.1

[a]The score values correspond to the difference between the maximum interaction likelihood of the peptides with domains of *Escherichia coli* and human ATP synthases.; [b]The charge values correspond to the balance between the number of acid (Glu and Asp) and basic (Lys and Arg) residues.; [c]The GRAVY index is provided as indicator of the hydrophilicity of the peptides [70].; [d]The helicity is obtained as the fraction of residues in alpha helix conformation, as predicted by PSIPred [71].; Bold indicates structural features that led to the selection of P1 and P2 as the top-ranked peptide inhibitors.

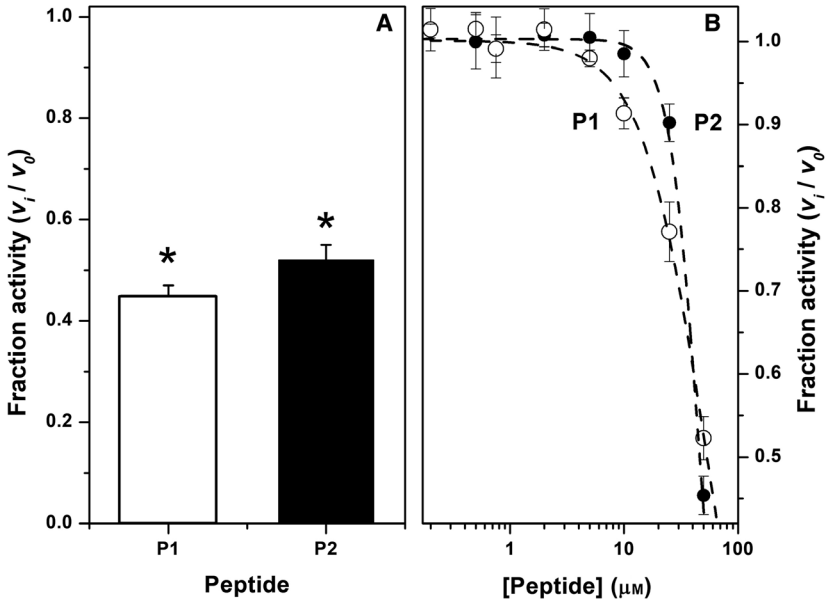

Fig. 3. Inhibition of ATPase activity of *Ec*F_1 by the designed peptides. v_0 and v_i are the initial catalytic velocities in absence and in presence of a given concentration of peptide [Peptide], respectively. (A) Fraction activity measured at 50 μM peptide concentration. Both peptides showed a statistically significant catalytic decrease in relation to the peptide-free enzyme (*$P < 0.01$ compared with the untreated control). The inhibitory potencies of P1 and P2 are also statistically different between each other at the 0.01 significance level. (B) Dose–response plot. The data were fitted using the Hill equation. Positive control measurements using quercetin as *Ec*F_1 inhibitor yielded $IC_{50} = 38 \pm 4$ μM and $h = 0.95 \pm 0.01$, which are in good agreement with published data [69]. All measurements were performed using the malachite green phosphate assay, with 1.1 μM *Ec*F_1, 1 mM ATP, 1 mM Mg(II), 50 mM Tris-SO_4, pH 8.0, 30 °C. Values are means ± SD from at least triplicates.

multiple interactions established through transient subunit–subunit interfaces [46–49]. The basic bioenergetic function of ATP synthase, complex as it is, is not the only function this enzyme is involved in. As previously mentioned, it has been found in many other locations of mammalian cells than mitochondria, playing an important role in regulating many human cellular processes. Consequently, ATP synthase has unexpectedly become a therapeutic drug target in the treatment of numerous diseases [50,51], besides its proven usefulness for antibiotic development [30]. After these findings, the inhibition of ATP synthase for clinical purposes has drawn special interest. A wide variety of molecules has been found to inhibit it, including exogenous natural compounds and peptides, and endogenous ATP synthase subunits [29,50]. This chemical spectrum of molecules inhibiting ATP synthase underscores the versatility of the enzyme as a molecular target and validates ongoing efforts by different groups to find and optimize new inhibitors targeting this enzyme to both develop new antibiotics or compounds to treat human disorders [29,52–54].

Although the F_OF_1-ATP synthase complex is inhibited by over 250 natural and synthetic inhibitors, the number of peptide-like inhibitors reported so far is relatively low [55]. Most of these peptides with promising pharmaceutical activities coming from exogenous natural sources such as venoms, sponges, and insects have been reported as inhibitors of *Ec*F_1 [56–59]. However, very few endogenous peptide inhibitors of the F_OF_1-ATP synthase complex have been discovered so far. IF_1, a small natural regulatory protein which binds to the hydrophilic F_1 portion [60,61], and the ϵ subunit of F_1 in bacteria and chloroplasts [62], are the most representative endogenous inhibitors of the ATP synthase. Nonetheless, the potentialities of other ATP synthase fragments for being tailored and turned into a novel inhibitor have not been considered in the *de novo* design of endogenous inhibitors. Here, we demonstrate the suitability of turning enzyme fragments into peptide inhibitors by combining the rational generation of diversity-oriented peptide libraries from such protein fragments and machine-learning algorithms evaluating the peptide–protein interactions.

De novo design of F_1-ATP synthase inhibitors

As mentioned in the introductory section, evolutionary algorithms have been hardly applied to the *in silico* generation of peptide libraries. In general, the

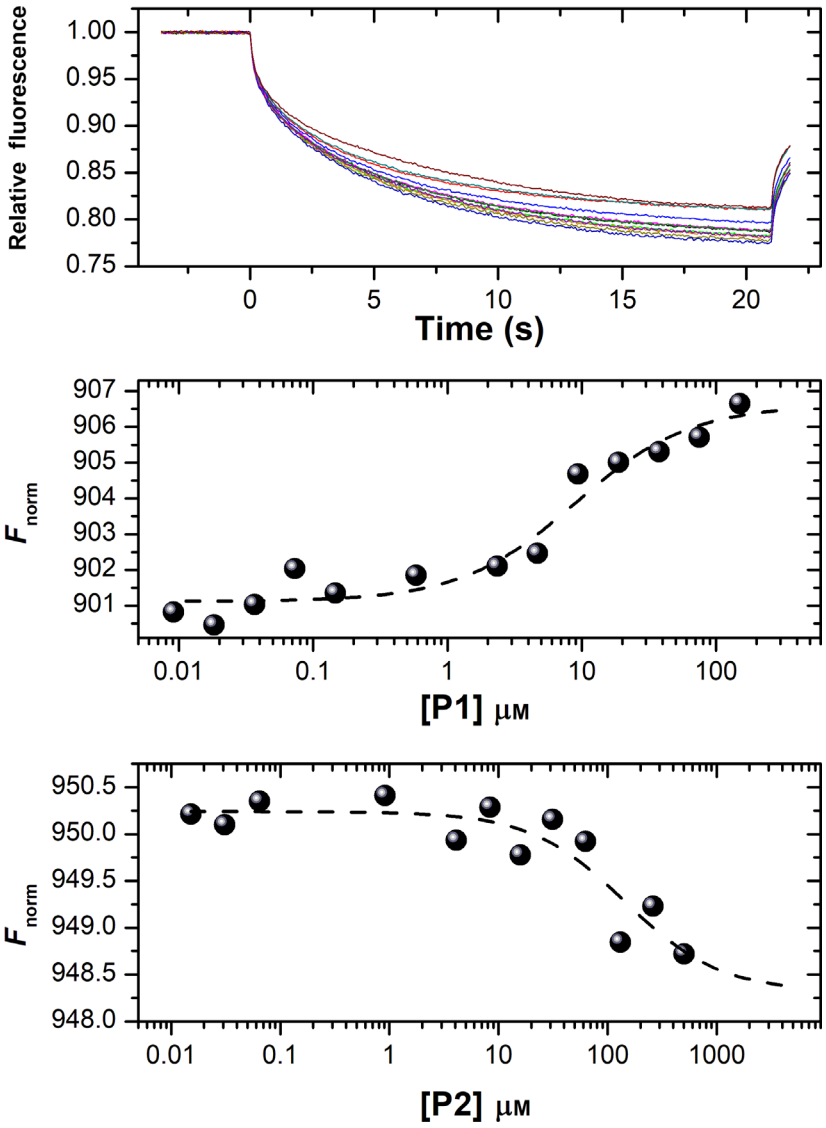

Fig. 4. Measurement of EcF1-peptide binding by MST. (A) MST traces for the titration of EcF1 with P1. (B) Titration curve with P1. (C) Titration curve with P2. All measurements were performed with 50 nM EcF1, 50 mM Tris-SO4, pH 8.0.

application of evolutionary approaches has been devoted to the optimization steps of peptide drugs. For instance, the simulated molecular evolution (SME) algorithm has recently been applied to improve the peptide selectivity for cancer cells after applying a machine-learning model predicting peptides with anticancer activity [13]. SME is a stochastic optimization algorithm that includes genetic variations on peptides/proteins showing a promising biological activity. This algorithm is tightly coupled to a selection process represented by artificial neural networks aimed to optimize a certain biological property [63,64].

Here, we propose a different approach for the design of bioactive peptides, which also leverages machine-learning models and evolutionary algorithms but in a different mode. Our strategy repurposes the simulation of sequences evolution to the rational generation of diversity-oriented peptide libraries that are subsequently explored with machine-learning models of PPI. This is achieved by applying a flexible evolutionary algorithm, as implemented in ROSE, that comprises parameters such as average genetic distance, tree topology, and insertion and deletion events, among others. The advantage of using evolutionary algorithms to build libraries of candidates lies in the application of previous knowledge on the sites/residues that account for biological activity when mutations are performed. Thus, a consensus (root) peptide with its corresponding conservation scoring profile can be used to assign different mutation rates to each position in the sequence. On the other hand, the sequence diversity of the peptides in the library can be controlled by

Table 2. Predicted interactions scores of P1 and P2 with subunits of the *Escherichia coli* and human FoF_1-ATP synthases.

Peptide	Subunit (range) [*E. coli*]	Score[a] [*E. coli*]	Subunit (range) [human]	Score[a] [human]
P1	**α(94–378)**	**0.52**	α(192–415)	0.38
	β (130–342)	**0.55**	β(186–405)	0.49
	γ (5–286)	**0.61**	γ (27–296)	0.46
	δ (7–175)	0.47	δ (38–162)	0.45
	ϵ (2–136)	0.43	ϵ	0.39
	a	0.28	a	0.27
	b	0.33	b	0.49
	c	0.24	c	0.25
P2	**α (94–378)**	**0.58**	α (137–418)	0.46
	β (130–342)	0.48	β (186–405)	0.37
	γ (5–286)	0.41	γ (27–296)	0.48
	δ (7–175)	0.38	δ (38–162)	0.38
	ϵ (2–136)	0.47	ϵ	0.43
	a	0.21	a	0.27
	b	0.34	b	0.28
	c	0.23	c	0.25

[a]Score values higher than 0.5 represent putative interactions.; Bold indicates interaction capabilities that led to the selection of P1 and P2 as the top-ranked peptide inhibitors.

evaluating the ROSE output with an all vs all global alignment [65]. Here, we calibrated ROSE parameters to produce peptide libraries with an overall 70% of identity by using our software called SeqDivA (Sequence Diversity Analysis), free-available at https://github.com/eancedeg/SeqDivA [66]. All these evolutionary considerations provide rationality to the generation of peptide libraries. Thus, the probability to find new biologically relevant peptides is higher than approaches using stochastic mutations. In the present study, the resulting peptide library was screened with PPI-Detect to identify putative binders of *E. coli* ATP synthase and, at the same time, to diminish the likelihood of binding with the human ATP synthase. Table 2 summarizes the obtained scores for the two identified inhibitors P1 and P2.

Figure 2 shows the locations of the native fragments related to P1 and P2 in a model ATP synthase structure. Notably, P1, which is derived from one of the chains in the stalk of the protein, is predicted to bind potentially α, β, and γ subunits (Table 2), that are all close or in direct contact with the original fragment in the protein. Similarly, P2 is predicted as a binder of β subunit, when precisely this peptide originates from a fragment of an α subunit, the natural partner of the β subunit in the hexamer ring of F_1-ATP synthase. These predictions are aligned with the base hypothesis of our design strategy; that is, the new peptides conserve the binding properties of their parents. In addition, the results prove the accuracy of PPI-Detect for identifying the right interaction partners among the peptides in the generated library.

The strategy described in this work constitutes the *de novo* design of ATP synthase inhibitors, using fragments of interfaces between subunits of the protein as starting points with no prior evidence of pharmacological activity. The two designed peptides displayed IC_{50}/EC_{50} values in the micromolar order, a potency that compares well with those reported for diverse exogenous peptide inhibitors of this enzyme [52,56]. Our immediate outlook involves improving the membrane permeability with derivatives of these peptides in order to evaluate their activity *in vivo* assays. Nonetheless, the success of the computational methodology used to design these peptides has been established and constitutes an attractive framework for the design of peptides to other enzymatic complexes or biomolecular systems, where the disruption of specific protein–protein interactions at the interface of subunits or domains can modulate their function.

Acknowledgements

YBR-B received a postdoctoral fellowship from DGAPA, Universidad Nacional Autónoma de México (UNAM). LPA-B and EH-G are, respectively, students from Programa de Doctorado and Maestría en Ciencias Bioquímicas, Universidad Nacional Autónoma de México (UNAM). LPA-B. received fellowship No. 275485 from CONACyT, México. This work was financed in part by DGAPA, UNAM [PAPIIT IN205018] and CONACyT [Grant 235831]. GA-C was supported by the Strategic Funding UID/Multi/04423/2019 through national funds provided by the Portuguese Foundation for Science and Technology (Fundação para a Ciência e a Tecnologia-FCT).

Author contributions

YBR-B, GA-C, and EG-H conceived and supervised the experiments involved in the presented strategy. YBR-B and GA-C performed the *in silico* analyses. LPA-B prepared reagents and performed *in vitro* assays. EH-G and LPA-B interpreted data. EH-G, LPA-B, and GA-C wrote the manuscript. YBR-B, AA, and EG-H made manuscript revisions. All authors have read and agreed to the published version of the manuscript.

References

1 Bernardini S, Tiezzi A, Laghezza Masci V and Ovidi E (2018) Natural products for human health: an historical

overview of the drug discovery approaches. *Nat Prod Res* **32**, 1926–1950.
2 Henninot A, Collins JC and Nuss JM (2018) The current state of peptide drug discovery: back to the future? *J Med Chem* **61**, 1382–1414.
3 Aaghaz S, Gohel V and Kamal A (2019) Peptides as potential anticancer agents. *Curr Top Med Chem* **19**, 1491–1511.
4 Shama N, Kumari R, Bais S and Shrivastava A (2019) Some important peptides and their role in obesity-current insights. *Obesity Medicine* **16**, 100144.
5 Hodgson WC and Isbister GK (2009) The application of toxins and venoms to cardiovascular drug discovery. *Curr Opin Pharmacol* **9**, 173–176.
6 Zhang LJ and Gallo RL (2016) Antimicrobial peptides. *Curr Biol* **26**, R14–R19.
7 Bozovicar K and Bratkovic T (2019) Evolving a peptide: library platforms and diversification strategies. *Int J Mol Sci* **21**, 1–31.
8 Wang YC and Distefano MD (2014) Synthesis and screening of peptide libraries with free C-termini. *Curr Top Pept Protein Res* **15**, 1–23.
9 Goodnow RA Jr, Dumelin CE and Keefe AD (2017) DNA-encoded chemistry: enabling the deeper sampling of chemical space. *Nat Rev Drug Discov* **16**, 131–147.
10 Marasco D, Perretta G, Sabatella M and Ruvo M (2008) Past and future perspectives of synthetic peptide libraries. *Curr Protein Pept Sci* **9**, 447–467.
11 Irving MB, Pan O and Scott JK (2001) Random-peptide libraries and antigen-fragment libraries for epitope mapping and the development of vaccines and diagnostics. *Curr Opin Chem Biol* **5**, 314–324.
12 Müller AT, Gabernet G, Hiss JA and Schneider G (2017) modlAMP: Python for antimicrobial peptides. *Bioinformatics* **33**, 2753–2755.
13 Gabernet G, Gautschi D, Müller AT, Neuhaus CS, Armbrecht L, Dittrich PS, Hiss JA and Schneider G (2019) In silico design and optimization of selective membranolytic anticancer peptides. *Sci Rep* **9**, 11282.
14 Stoye J, Evers D and Meyer F (1998) Rose: generating sequence families. *Bioinformatics* **14**, 157–163.
15 Pang A, Smith AD, Nuin PAS and Tillier ERM (2005) SIMPROT: using an empirically determined indel distribution in simulations of protein evolution. *BMC Bioinformatics* **6**, 236.
16 Fletcher W and Yang Z (2009) INDELible: a flexible simulator of biological sequence evolution. *Mol Biol Evol* **26**, 1879–1888.
17 Romero-Molina S, Ruiz-Blanco YB, Harms M, Münch J and Sanchez-Garcia E (2019) PPI-Detect: a support vector machine model for sequence-based prediction of protein-protein interactions. *J Comput Chem* **40**, 1233–1242.
18 Ding Z and Kihara D (2018) Computational methods for predicting protein-protein interactions using various protein features. *Curr Protoc Protein Sci* **93**, e62.
19 Savojardo C, Martelli PL and Casadio R (2020) Protein-protein interaction methods and protein phase separation. *Annu Rev Biomed Data Sci* **3**, 89–112.
20 Park Y (2009) Critical assessment of sequence-based protein-protein interaction prediction methods that do not require homologous protein sequences. *BMC Bioinformatics* **10**, 419.
21 Ahmad Z and Laughlin TF (2010) Medicinal chemistry of ATP synthase: a potential drug target of dietary polyphenols and amphibian antimicrobial peptides. *Curr Med Chem* **17**, 2822–2836.
22 Ahmad Z, Okafor F, Azim S and Laughlin TF (2013) ATP synthase: a molecular therapeutic drug target for antimicrobial and antitumor peptides. *Curr Med Chem* **20**, 1956–1973.
23 Boyer PD (1997) The ATP synthase—a splendid molecular machine. *Annu Rev Biochem* **66**, 717–749.
24 Cingolani G and Duncan TM (2011) Structure of the ATP synthase catalytic complex (F(1)) from *Escherichia coli* in an autoinhibited conformation. *Nat Struct Mol Biol* **18**, 701–707.
25 Sobti M, Smits C, Wong ASW, Ishmukhametov R, Stock D, Sandin S and Stewart AG (2016) Cryo-EM structures of the autoinhibited *E. coli* ATP synthase in three rotational states. *Elife* **5**, 1–18.
26 Yoshida M, Muneyuki E and Hisabori T (2001) ATP synthase–a marvellous rotary engine of the cell. *Nat Rev Mol Cell Biol* **2**, 669–677.
27 Chi SL and Pizzo SV (2006) Cell surface F1Fo ATP synthase: a new paradigm? *Ann Med* **38**, 429–438.
28 Kucharczyk R, Zick M, Bietenhader M, Rak M, Couplan E, Blondel M, Caubet S-D and di Rago J-P (2009) Mitochondrial ATP synthase disorders: molecular mechanisms and the quest for curative therapeutic approaches. *Biochim Biophys Acta* **1793**, 186–199.
29 Nesci S, Trombetti F, Algieri C and Pagliarani A (2019) A therapeutic role for the F1FO-ATP synthase. *SLAS Discov* **24**, 893–903.
30 Andries K, Verhasselt P, Guillemont J, Göhlmann HW, Neefs JM, Winkler H, Van Gestel J, Timmerman P, Zhu M, Lee E *et al.* (2005) A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*. *Science* **307**, 223–227.
31 Perdih A, Hrast M, Barreteau H, Gobec S, Wolber G and Solmajer T (2014) Inhibitor design strategy based on an enzyme structural flexibility: a case of bacterial MurD ligase. *J Chem Inf Model* **54**, 1451–1466.
32 Nekrasov AN and Zinchenko AA (2010) Structural features of the interfaces in enzyme-inhibitor complexes. *J Biomol Struct Dyn* **28**, 85–96.
33 Gouy M, Guindon S and Gascuel O (2010) SeaView version 4: a multiplatform graphical user interface for

sequence alignment and phylogenetic tree building. *Mol Biol Evol* **27**, 221–224.
34 Romero-Molina S, Ruiz-Blanco YB, Green JR and Sanchez-Garcia E (2019) ProtDCal-Suite: a web server for the numerical codification and functional analysis of proteins. *Protein Sci* **28**, 1734–1743.
35 Ruiz-Blanco YB, Paz W, Green J and Marrero-Ponce Y (2015) ProtDCal: a program to compute general-purpose-numerical descriptors for sequences and 3D-structures of proteins. *BMC Bioinformatics* **16**, 162.
36 Mosca R, Céol A, Stein A, Olivella R and Aloy P (2014) 3did: a catalog of domain-based interactions of known three-dimensional structure. *Nucleic Acids Res* **42** (Database issue), D374–D379.
37 Finn RD, Marshall M and Bateman A (2005) iPfam: visualization of protein-protein interactions in PDB at domain and amino acid resolutions. *Bioinformatics* **21**, 410–412.
38 Blohm P, Frishman G, Smialowski P, Goebels F, Wachinger B, Ruepp A and Frishman D (2014) Negatome 2.0: a database of non-interacting proteins derived by literature mining, manual annotation and protein structure analysis. *Nucleic Acids Res* **42** (Database issue), D396–D400.
39 Weber J, Lee RS, Grell E, Wise JG and Senior AE (1992) On the location and function of tyrosine beta 331 in the catalytic site of *Escherichia coli* F1-ATPase. *J Biol Chem* **267**, 1712–1718.
40 Duncan T and Senior A (1985) The defective proton-ATPase of uncD mutants of *Escherichia coli*. Two mutations which affect the catalytic mechanism. *J Biol Chem* **260**, 4901–4907.
41 Li Y, Ma X and Weber J (2019) Interaction between γC87 and γR242 residues participates in energy coupling between catalysis and proton translocation in *Escherichia coli* ATP synthase. *Biochim Biophys Acta Bioenerg* **1860**, 679–687.
42 Van Veldhoven PP and Mannaerts GP (1987) Inorganic and organic phosphate measurements in the nanomolar range. *Anal Biochem* **161**, 45–48.
43 Jerabek-Willemsen M, André T, Wanner R, Roth HM, Duhr S, Baaske P and Breitsprecher D (2014) MicroScale thermophoresis: interaction analysis and beyond. *J Mol Struct* **1077**, 101–113.
44 Huang Y, Niu B, Gao Y, Fu L and Li W (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **26**, 680–682.
45 Aguilera-Mendoza L, Marrero-Ponce Y, Tellez-Ibarra R, Llorente-Quesada MT, Salgado J, Barigye SJ and Liu J (2015) Overlap and diversity in antimicrobial peptide databases: compiling a non-redundant set of sequences. *Bioinformatics* **31**, 2553–2559.
46 Abrahams JP, Leslie AG, Lutter R and Walker JE (1994) Structure at 2.8 A resolution of F1-ATPase from bovine heart mitochondria. *Nature* **370**, 621–628.
47 Pulido NO, Salcedo G, Pérez-Hernández G, José-Núñez C, Velázquez-Campoy A and García-Hernández E (2010) Energetic effects of magnesium in the recognition of adenosine nucleotides by the F(1)-ATPase beta subunit. *Biochemistry* **49**, 5258–5268.
48 Hahn-Herrera O, Salcedo G, Barril X and García-Hernández E (2016) Inherent conformational flexibility of F1-ATPase alpha-subunit. *Biochim Biophys Acta* **1857**, 1392–1402.
49 Mendoza-Hoffmann F, Zarco-Zavala M, Ortega R and García-Trejo JJ (2018) Control of rotation of the F1FO-ATP synthase nanomotor by an inhibitory alpha-helix from unfolded epsilon or intrinsically disordered zeta and IF1 proteins. *J Bioenerg Biomembr* **50**, 403–424.
50 Hong S and Pedersen PL (2008) ATP synthase and the actions of inhibitors utilized to study its roles in human health, disease, and other scientific areas. *Microbiol Mol Biol Rev* **72**, 590–641. Table of Contents.
51 Johnson JA and Ogbi M (2011) Targeting the F1Fo ATP synthase: modulation of the body's powerhouse and its implications for human disease. *Curr Med Chem* **18**, 4684–4714.
52 Hotra A, Ragunathan P, Ng PS, Seankongsuk P, Harikishore A, Sarathy JP, Saw W-G, Lakshmanan U, Sae-Lao P, Kalia NP *et al.* (2020) Discovery of a novel mycobacterial F-ATP synthase inhibitor and its potency in combination with diarylquinolines. *Angew Chem Int Ed Engl* **59**, 13295–13304.
53 Chinnam N, Dadi PK, Sabri SA, Ahmad M, Kabir MA and Ahmad Z (2010) Dietary bioflavonoids inhibit *Escherichia coli* ATP synthase in a differential manner. *Int J Biol Macromol* **46**, 478–486.
54 Yarlagadda V, Medina R and Wright GD (2020) Venturicidin A, a membrane-active natural product inhibitor of ATP synthase potentiates aminoglycoside antibiotics. *Sci Rep* **10**, 8134.
55 Narang R, Kumar R, Kalra S, Nayak SK, Khatik GL, Kumar GN, Sudhakar K and Singh SK (2019) Recent advancements in mechanistic studies and structure activity relationship of FoF1 ATP synthase inhibitor as antimicrobial agent. *Eur J Med Chem* **182**, 111644.
56 Syed H, Tauseef M and Ahmad Z (2018) A connection between antimicrobial properties of venom peptides and microbial ATP synthase. *Int J Biol Macromol* **119**, 23–31.
57 Ahmad Z, Tayou J and Laughlin TF (2015) Asp residues of betaDELSEED-motif are required for peptide binding in the *Escherichia coli* ATP synthase. *Int J Biol Macromol* **75**, 37–43.
58 Laughlin TF and Ahmad Z (2010) Inhibition of *Escherichia coli* ATP synthase by amphibian antimicrobial peptides. *Int J Biol Macromol* **46**, 367–374.
59 Azim S, McDowell D, Cartagena A, Rodriguez R, Laughlin TF and Ahmad Z (2016) Venom peptides

cathelicidin and lycotoxin cause strong inhibition of *Escherichia coli* ATP synthase. *Int J Biol Macromol* **87**, 246–251.
60 Pullman ME and Monroy GC (1963) A naturally occurring inhibitor of mitochondrial adenosine triphosphatase. *J Biol Chem* **238**, 3762–3769.
61 Nakamura J, Fujikawa M and Yoshida M (2013) IF1, a natural inhibitor of mitochondrial ATP synthase, is not essential for the normal growth and breeding of mice. *Biosci Rep* **33**, 735–741.
62 Kato-Yamada Y, Bald D, Koike M, Motohashi K, Hisabori T and Yoshida M (1999) Epsilon subunit, an endogenous inhibitor of bacterial F(1)-ATPase, also inhibits F(0)F(1)-ATPase. *J Biol Chem* **274**, 33991–33994.
63 Schneider G and Wrede P (1994) The rational design of amino acid sequences by artificial neural networks and simulated molecular evolution: de novo design of an idealized leader peptidase cleavage site. *Biophys J* **66** (2 Pt 1), 335–344.
64 Schneider G, Schrödl W, Wallukat G, Müller J, Nissen E, Rönspeck W, Wrede P and Kunze R (1998) Peptide design by artificial neural networks and computer-based evolutionary search. *Proc Natl Acad Sci USA* **95**, 12179–12184.
65 Needleman SB and Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* **48**, 443–453.
66 Aguero-Chapin G, Galpert D, Molina-Ruiz R, Ancede-Gallardo E, Pérez-Machado G, de la Riva GA and Antunes A (2019) Graph theory-based sequence descriptors as remote homology predictors. *Biomolecules* **10**, 1–35.
67 Rees DM, Leslie AG and Walker JE (2009) The structure of the membrane extrinsic region of bovine ATP synthase. *Proc Natl Acad Sci USA* **106**, 21597–21601.
68 Orriss GL, Leslie AG, Braig K and Walker JE (1998) Bovine F1-ATPase covalently inhibited with 4-chloro-7-nitrobenzofurazan: the structure provides further support for a rotary catalytic mechanism. *Structure* **6**, 831–837.
69 Dadi PK, Ahmad M and Ahmad Z (2009) Inhibition of ATPase activity of *Escherichia coli* ATP synthase by polyphenols. *Int J Biol Macromol* **45**, 72–79.
70 Kyte J and Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105–132.
71 Buchan DWA and Jones DT (2019) The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res* **47**, W402–W407.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Schematic description of the ROSE algorithm.
Table S1. Consensus peptides derived from the MSAs.
File S1. Multiple Sequence Alignments using MAFFT for the ATP synthase interfaces (Class 1-4).
File S2. Input file for ROSE for the generation of related peptides from a consensus peptide (root).