Check for updates

Received: 19 June 2023 | Accepted: 6 October 2023
DOI: 10.1002/prot.26619

PROTEINS WILEY

RESEARCH ARTICLE

Scaffold Matcher: A CMA-ES based algorithm for identifying hotspot aligned peptidomimetic scaffolds

Erin R. Claussen[1] | P. Douglas Renfrew[2] | Christian L. Müller[3,4,5] | Kevin Drew[1]

[1]Department of Biological Sciences, University of Illinois at Chicago, Chicago, Illinois, USA
[2]Center for Computational Biology, Flatiron Institute, New York, New York, USA
[3]Department of Statistics, Ludwig-Maximilians-Universität München, Munich, Germany
[4]Institute of Computational Biology, Helmholtz Zentrum München, Munich, Germany
[5]Center for Computational Mathematics, Flatiron Institute, New York, New York, USA

Correspondence
Kevin Drew, Department of Biological Sciences, University of Illinois at Chicago, Chicago, IL 60607, USA.
Email: ksdrew@uic.edu

Funding information
National Institutes of Health, Grant/Award Numbers: L40 HD096554, R00 HD092613; RosettaCommons, Grant/Award Number: RC22024

Abstract

The design of protein interaction inhibitors is a promising approach to address aberrant protein interactions that cause disease. One strategy in designing inhibitors is to use peptidomimetic scaffolds that mimic the natural interaction interface. A central challenge in using peptidomimetics as protein interaction inhibitors, however, is determining how best the molecular scaffold aligns to the residues of the interface it is attempting to mimic. Here we present the Scaffold Matcher algorithm that aligns a given molecular scaffold onto hotspot residues from a protein interaction interface. To optimize the degrees of freedom of the molecular scaffold we implement the covariance matrix adaptation evolution strategy (CMA-ES), a state-of-the-art derivative-free optimization algorithm in Rosetta. To evaluate the performance of the CMA-ES, we used 26 peptides from the FlexPepDock Benchmark and compared with three other algorithms in Rosetta, specifically, Rosetta's default minimizer, a Monte Carlo protocol of small backbone perturbations, and a Genetic algorithm. We test the algorithms' performance on their ability to align a molecular scaffold to a series of hotspot residues (i.e., constraints) along native peptides. Of the 4 methods, CMA-ES was able to find the lowest energy conformation for all 26 benchmark peptides. Additionally, as a proof of concept, we apply the Scaffold Match algorithm with CMA-ES to align a peptidomimetic oligooxopiperazine scaffold to the hotspot residues of the substrate of the main protease of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Our implementation of CMA-ES into Rosetta allows for an alternative optimization method to be used on macromolecular modeling problems with rough energy landscapes. Finally, our Scaffold Matcher algorithm allows for the identification of initial conformations of interaction inhibitors that can be further designed and optimized as high-affinity reagents.

KEYWORDS
CMA-ES, covariance matrix adaptation evolution strategy, derivative-free optimization, hotspot residues, peptidomimetic, Rosetta, scaffold matcher

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2023 The Authors. *Proteins: Structure, Function, and Bioinformatics* published by Wiley Periodicals LLC.

1 | INTRODUCTION

Protein interactions govern a wide range of biological functions. Aberrant protein interactions disrupt cellular function and often lead to human disease.[1] The design of protein interaction inhibitors is one strategy to correct aberrant protein interactions.[2] A promising approach to designing protein interaction inhibitors is to mimic the residues at the interface of one binding partner using a stable and synthetically tractable molecular scaffold known as a peptidomimetic.[3] Residues at the interface of a protein interaction that contribute most to the binding affinity are designated hotspot residues[4] and are ideal residues to mimic when designing inhibitors. A central challenge in designing peptidomimetic inhibitors is the identification of a peptidomimetic molecular scaffold that aligns with the hotspot residues at the interface of the interaction. Unfortunately, there are limited tools available in which to achieve this goal. However, recent advances in molecular modeling allow for this challenge to be addressed computationally.[5–7]

The Rosetta macromolecular modeling toolkit is a computational package that offers users and developers state-of-the-art methods for important applications such as structure prediction, protein docking, loop modeling, protein design, antibody modeling, and many others.[5,8] Noncanonical backbones (i.e., peptidomimetics) have been incorporated into Rosetta to allow for the modeling and design of scaffolds such as oligooxopiperazines, peptoids, stapled helices, among others.[6,9] We have previously used Rosetta to design oligooxopiperazines inhibitors of MDM2 mimicking P53 hotspot residues as well as inhibitors of P300 mimicking HIF1alpha hotspot residues.[7] However, in these cases, the starting placement of the oligooxopiperazine scaffold was aligned manually to the Cß atoms of the hotspot residues without considering the scaffold internal degrees of freedom or evaluating the energy of the molecule until later downstream analysis. This leaves open the question of whether there are more optimal conformations of the scaffold aligned with the hotspots.

We therefore introduce the Scaffold Matcher algorithm in Rosetta. The Scaffold Matcher algorithm aligns a given molecular scaffold (e.g., peptidomimetic) onto hotspot residues from a protein binding interface and scores the alignment given Rosetta's energy function. To efficiently traverse the Rosetta's energy function, we implemented the covariance matrix adaptation evolution strategy (CMA-ES) algorithm. CMA-ES is a stochastic gradient-free method for optimizing nonconvex functions.[10] CMA-ES starts by taking many samples from an isotropic multivariate normal distribution that are distributed broadly across the energy landscape rather than centered in one local energy minimum. The algorithm also takes advantage of updating the covariance matrix which increases the probability of optimal search *directions*. Originally proposed as an alternative to conjugate gradient methods,[11] CMA-ES has been shown to outperform many other optimization algorithms on standard benchmarks[12] and has recently been applied to structural modeling problems with promising results.[13–19]

While the ultimate purpose of the Scaffold Matcher algorithm is to align peptidomimetic scaffolds to hotspot residues, there are no curated benchmarks of peptidomimetics bound to their protein partners. To evaluate our algorithm's performance on a gold standard test set, we therefore use the FlexPepDock benchmark of 26 peptides, and compare CMA-ES's results to three other algorithms, a Genetic algorithm, Rosetta's gradient descent minimizer, and a Monte Carlo protocol of small backbone perturbations. We show CMA-ES's performance is superior using multiple metrics of structural comparison and show a time analysis demonstrating CMA-ES is competitive or even superior with other algorithms. Finally, we show a proof of principle demonstration of the Scaffold Matcher algorithm's use on matching hotspot residues of the SARS-CoV-2 main protease (mPro) substrate peptide using an oligooxopiperazine peptidomimetic scaffold.

2 | RESULTS

Figure 1 shows a general scheme for alignment of a scaffold onto hotspot residues. First, a protein complex is identified (Figure 1A), and one subunit of the complex is selected to be the target while the other is selected to be mimicked. Hotspot residues are determined from the mimicked subunit and are isolated in space by removing the backbone atoms (Figure 1B). A peptidomimetic scaffold is then aligned onto the hotspot residue positions therefore mimicking their role at the protein interaction interface (Figure 1C). Many peptidomimetic scaffolds are similar to peptides and therefore have similar degrees of freedom allowing flexibility in their structure (Figure 1D). To address the challenge of matching a molecular scaffold to a set of hotspot residues, we first require an optimization algorithm that will efficiently sample these degrees of freedom of the molecular scaffold. We therefore implemented CMA-ES, an evolutionary optimization algorithm, into the Rosetta framework.

2.1 | CMA-ES algorithm

CMA-ES generates samples from a multivariate normal distribution where the samples are then ranked using the objective function to be minimized. Based on this ranking, a subset of the samples is then selected and used to update the mean and covariance matrix of the multivariate normal distribution for the next iteration. The process is repeated until convergence. In principle, CMA-ES can be regarded as an iterative (biased) principal component analysis on previously selected samples to adapt future sampling directions. We fully detail the mathematical description of the algorithm in Supplementary Methods (File S1) and include a graphical representation of the algorithm's procedure in Figure S1.

2.2 | CMA-ES algorithm evaluation on model tripeptide

The CMA-ES algorithm is a general purpose optimization algorithm, which can be utilized independently of matching molecular scaffolds

10970134, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/prot.26619 by Helmholtz Zentrum Muenchen Deutsches Forschungszentrum, Wiley Online Library on [07/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

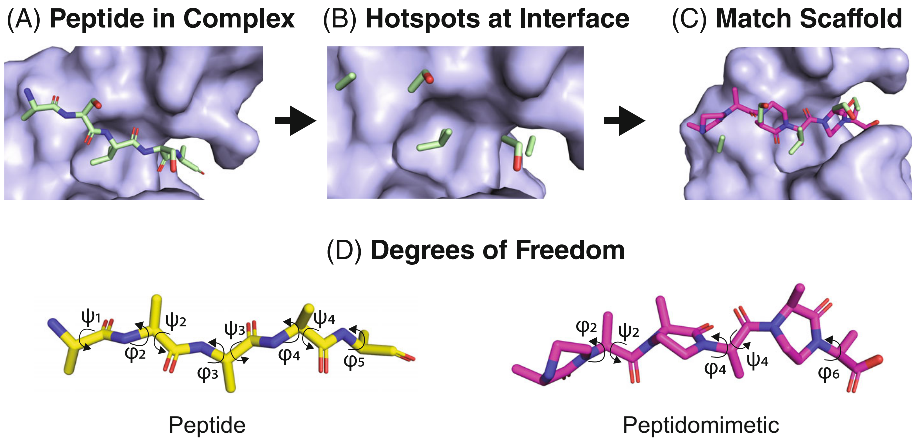

FIGURE 1 Identification of hotspot mimetic scaffolds for interaction inhibitors. (A) Peptide to be mimicked is identified in complex with partner protein. (B) Hotspot residues at the interface are identified and sidechain coordinates recorded. (C) Identify alternative molecular scaffolds (i.e., peptidomimetics) that mimic the placement of the hotspot residues. (D) Primary degrees of freedom for peptides and oligooxopiperazines used during optimization procedure to fit scaffold to hotspot residues. Omega torsion angles and ring constrained phi/psi torsion angles not labeled for clarity.

to hotspot residues. We therefore first demonstrate our implementation of the CMA-ES algorithm in Rosetta by minimizing a model tri-peptide (i.e., Ala—His—Ala) from a high energy state to a lower energy state. The CMA-ES algorithm generates samples from a multivariate normal distribution. In the case of peptide minimization, the algorithm samples dihedral angles of the peptide backbone and side chains to generate a new conformation (i.e., pose). Once new conformations are generated, each resulting pose is scored using the Rosetta energy function. The poses are ranked according to their score. Top scoring conformations are used to adapt the sampling distribution for the generation of new poses. Figure S2A shows conformations from selected iterations throughout the algorithm's procedure when applied to the tri-peptide. We observe a substantial amount of conformational sampling by iteration 30 allowing the algorithm to explore a rough energy landscape and escape local energy minima. By iterations 60 and 90 sampling focuses on exploring a more local area of the conformational space. And finally, by iteration 120 the algorithm converges toward sampling around the (potentially) global minimum conformation.

We can also observe how individual dihedral angles are being sampled throughout the optimization process. Unlike a Monte Carlo simulation, for example, which is sampling from an a priori fixed distribution, the CMA-ES algorithm is updating the multivariate distribution every iteration. Therefore, individual dihedral angles may be sampled variably during optimization and in a correlated manner. We can observe this behavior in Figure S2B. We see dihedral angles along the backbone and Histidine side chain of the tri-peptide dynamically change their values as a function of CMA-ES iteration. In particular, we see the first Alanine residue's Psi angle being sampled continuously through iteration 90, whereas the Chi angles of the Histidine side chain are relatively constant throughout optimization. This shows that the CMA-ES algorithm efficiently samples only the degrees of freedom necessary to find a low energy conformation.

A major feature of the CMA-ES algorithm is its ability to navigate rough energy landscapes. Consistent with our observations above where we see a large amount of conformational sampling (Figure S2A) we also see increases in Rosetta energy at various iterations along the optimization procedure (Figure S2C). This increase in conformational sampling and energy allows for the escape out of local energy minima. In particular, we observe a spike in Rosetta energy at iteration 30 where the largest amount of conformational sampling is seen. Rosetta energy is further seen to decrease with less variance until convergence.

2.3 | Scaffold Matcher algorithm

We next introduce the Scaffold Matcher algorithm, designed to identify a low-energy conformation of a given molecular scaffold aligned to a set of hotspot residues. The algorithm can accommodate a range of molecular scaffolds including oligooxopiperazines and peptides as shown in Figure 1D. Figure 2 shows the overall workflow of the Scaffold Matcher algorithm. As shown in Figure 2A, the first step of the workflow begins with a target peptide bound to a protein. The peptide is extracted from the target keeping the peptide's conformation fixed (Figure 2B). Next, backbone atoms of the peptide are removed from the pose, leaving only the sidechain atoms (i.e., Disembodied Sidechains, Figure 2C). We then select one sidechain residue to be the primary hotspot and all other side-chain residues as ancillary hotspots. The primary hotspot is often selected based on the value from a ΔΔG calculation or some other prior information suggesting the residue is important for binding.

The protocol next puts energy constraints between the atoms of the disembodied side-chain residues and the corresponding residues on the input molecular scaffold. Constraints are used as described in Fleishman et al.[20] and are mathematically described in the

10970134, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/prot.26619 by Helmholtz Zentrum Muenchen Deutsches Forschungszentrum, Wiley Online Library on [07/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

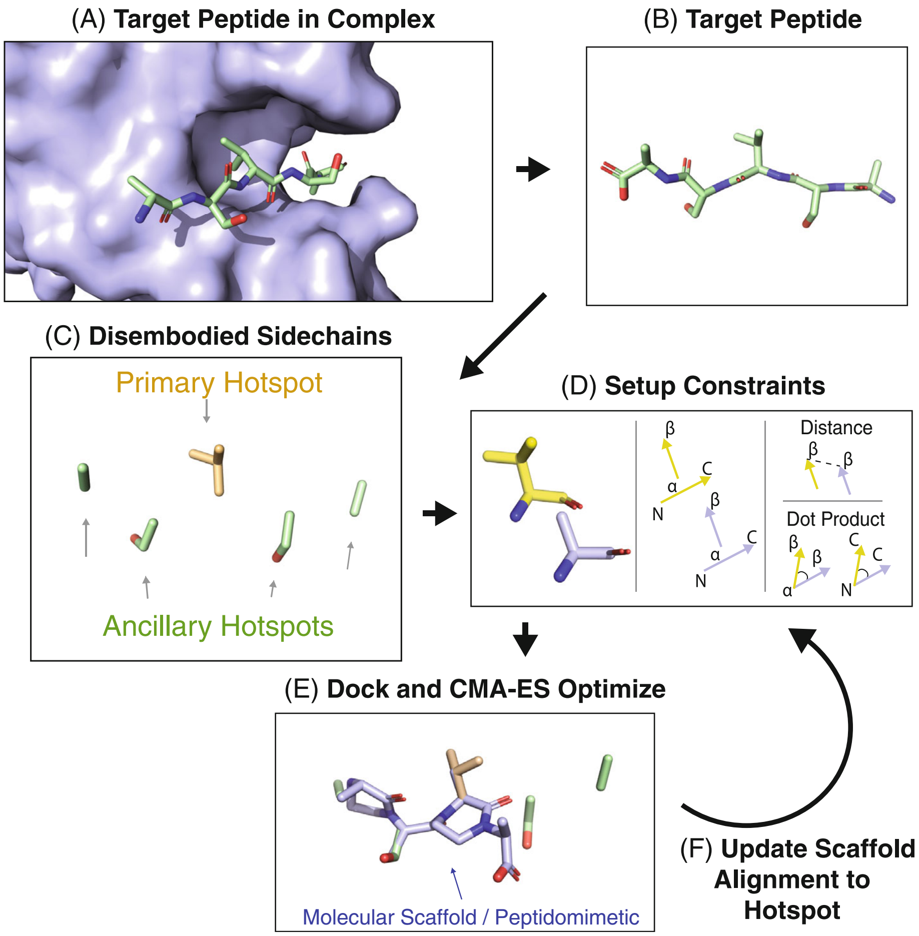
(A) Target Peptide in Complex
(B) Target Peptide
(C) Disembodied Sidechains
Primary Hotspot
Ancillary Hotspots
(D) Setup Constraints
β
C
α
N
β
α
C
N
Distance
β
β
Dot Product
β
C
β
C
α
N
(E) Dock and CMA-ES Optimize
Molecular Scaffold / Peptidomimetic
(F) Update Scaffold Alignment to Hotspot

FIGURE 2 Scaffold Matcher algorithm overview. The figure displays a general overview of the Scaffold Matcher workflow which begins with a peptide (sticks) docked into a protein (surface blue-gray) (A). The target peptide is first isolated from the protein (B). Next, backbone atoms are removed from hotspot residues of the target peptide leaving only side chain atoms called disembodied side chains (C). One side chain is selected as the primary hotspot and all others are labeled as ancillary hotspots. Next, constraints are added between residues on an idealized molecular scaffold (e.g., peptidomimetic) and the hotspot residues (D). The molecular scaffold is then aligned onto the primary hotspot and phi and psi angles are optimized to match the idealized scaffold residue positions with the ancillary hotspots minimizing the constraints (E). The optimization is done using the covariance matrix adaptation evolution strategy (CMA-ES) optimization protocol where CMA-ES samples each degree of freedom updating the mean and standard deviation of the multivariate normal distribution every iteration. The process is iterated for new alignments of the molecular scaffold and the primary hotspot residue (F).

Supplementary Methods (File S1). Each constraint is calculated as a measure of the overlap between the Cβs, the Cα-Cβ vectors, and the C—N vectors of the disembodied residue and the idealized scaffold residue (Figure 2D). The constraint is added to the Rosetta energy function and if fully satisfied will result in a minimum −3 Rosetta Energy Units (REU) per residue. The molecular scaffold is then positioned to align position *i* on the scaffold to the primary hotspot residue. The remaining task is to identify phi and psi angle values that align the backbone to the disembodied residues and satisfy the energy constraints. To accomplish this, we utilize our implementation of the CMA-ES optimization algorithm in the Rosetta framework described above. Once a low-energy conformation is identified for this given alignment, a new alignment is created between the scaffold's position *i* + 1 and the primary hotspot residue. See Section 3 and Supplementary Methods (File S1) for full description of the Scaffold Matcher algorithm.

2.4 | Comparison of algorithms on individual examples

Due to the limited number of examples of peptidomimetics bound to target proteins, we chose to evaluate the Scaffold Matcher algorithm using peptides as a molecular scaffold. We applied the Scaffold Matcher to a set of 26 peptides in complex with proteins from the FlexPepDock benchmark.[21] For this analysis, all residues on the peptides were considered hotspots and the primary hotspot was chosen arbitrarily. Further, the peptide molecular scaffold was constructed of the same length and secondary structure of the target peptide with idealized phi/psi angles. Figure 3 highlights individual examples of peptides. In addition to using the CMA-ES optimization algorithm, we also benchmarked Rosetta's default gradient-based minimizer, a Genetic algorithm, and a Monte Carlo algorithm (see Section 3). To compare the four algorithms' performance, we calculated RMSD to the native structurally determined

10970134, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/prot.26619 by Helmholtz Zentrum Muenchen Deutsches Forschungszentrum, Wiley Online Library on [07/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

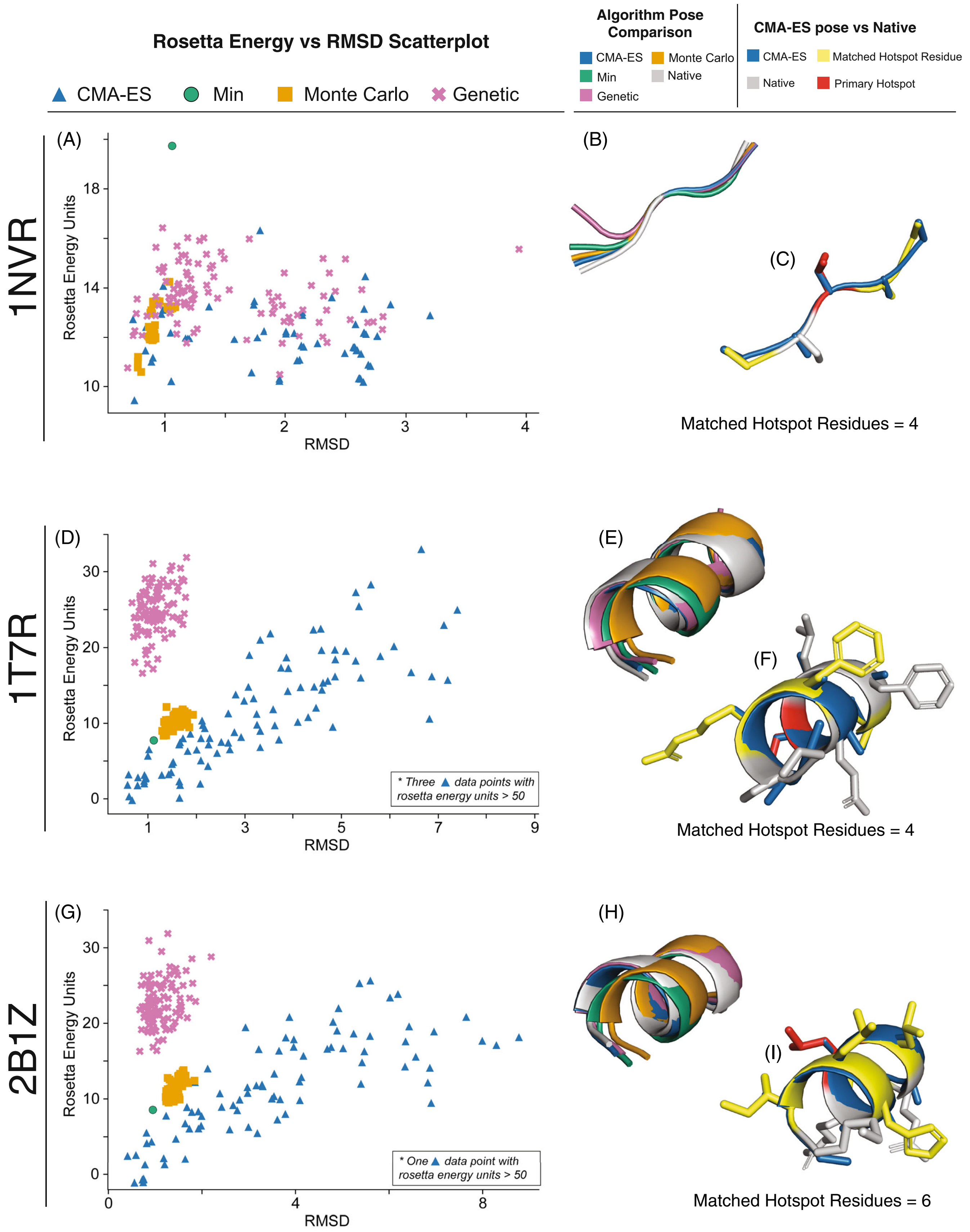

FIGURE 3 Legend on next page.

10970134, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/prot.26619 by Helmholtz Zentrum Muenchen Deutsches Forschungszentrum, Wiley Online Library on [07/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

peptide in addition to calculating the Rosetta energy score for the lowest energy pose of each individual trajectory and plotted a Rosetta Energy versus RMSD scatterplot. We observe from the example in PDBID 1NVR in Figure 3A that CMA-ES (blue triangle) outperforms Minimizer (green circle), Monte Carlo (orange square), and Genetic (pink x) sampling algorithms, having lower Rosetta energy. CMA-ES also produced a pose with lower RMSD to native compared with Monte Carlo and Minimizer. When we specifically compare to the Minimizer algorithm, the top CMA-ES model outperforms the Minimizer model by >10 Rosetta Energy Units (REU) and >0.3 RMSD. We next inspect the three-dimensional coordinates of the poses to the native and observe good alignment of the top CMA-ES pose with the native peptide (Figure 3B). Further, in Figure 3C, we see four out of the five hotspot residue constraints are matched with the CMA-ES algorithm. In comparison, only three hotspot residue constraints are matched with Genetic, two are matched with Monte Carlo, and one (the primary hotspot) is matched using the Minimizer (see Table S1).

Similar to 1NVR, we see CMA-ES outperform the Minimizer and Monte Carlo algorithms on other examples as well. In Figure 3D,G CMA-ES shows superior performance versus the other algorithms, improving on both energy (REU) and RMSD for peptides in PDBs 1T7R and 2B1Z respectively. When we inspect the poses overlaid with the native peptides (Figure 3E,H), we see a substantial improvement in terms of RMSD. The CMA-ES poses (blue) match very well to the native (gray) while the Minimizer (green) and Monte Carlo (orange) produce poses that deviate from the native. We also identify four matched hotspot residues by CMA-ES for 1T7R and six matched hotspot residues by CMA-ES for 2B1Z (Figure 3F,I). This is better than the Minimizer and Monte Carlo algorithms which had at most three matched hotspot residues but often lower (see Table S1). Additional benchmark examples can be found in Figure S3 which shows CMA-ES improved performance over the other algorithms.

2.5 | Global comparison of algorithms

Now that we have shown individual examples of CMA-ES outperforming other optimization algorithms, we evaluate each algorithm's models on all 26 peptides in the benchmark for a global comparison. Figure 4A shows a bar graph of the number of benchmark instances where each algorithm matched more hotspot residues than the other algorithms. In this winner-take-all analysis, CMA-ES matched more hotspot residues than the other algorithms for 11 benchmark instances while the Genetic algorithm and Monte Carlo produced the top model for only two and one benchmark instances, respectively. Minimizer did not match the most hotspot residues for any benchmark instance. Alternatively, we evaluate the difference in the number of matched hotspot residues between all four algorithms. This analysis provides a quantitative look at not just which algorithm had the top scoring model but by how much. In Figure S4A,C, we observe many benchmark instances where the CMA-ES algorithm matched +1, +2, +3, and +4 hotspot residues relative to the Monte Carlo and Genetic algorithms while there were no instances where the Monte Carlo or Genetic algorithm matched greater than +1 compared with CMA-ES. Similar behavior can be observed in Figure S4B, where many instances of CMA-ES match > +2 compared with the Minimizer algorithm while no instances were observed where the Minimizer matched more than +2 in comparison to CMA-ES. We also compare the other algorithms (Figure S4D–F) and observe the Monte Carlo and Genetic algorithms perform better than Minimizer. The Monte Carlo and Genetic algorithms are comparable in number of matched hotspots.

In addition to matching hotspot residues, we also evaluate each algorithms' performance in regards to total Rosetta energy. Figure 4B shows the CMA-ES models scored lower energies for all benchmark instances. Figure S4G–I shows the difference in Rosetta energy between models produced by CMA-ES and other algorithms. The mean difference between Rosetta energies of CMA-ES and Monte Carlo algorithms is −4.9 REU but as Figure S4G shows, for some benchmark instances, the CMA-ES models scored substantially lower in energy compared with Monte Carlo models (~10 REU). When we compare CMA-ES to the Minimizer algorithm, we see a more pronounced effect (Figure S4H) with CMA-ES models having substantially lower Rosetta energies including a mean difference in Rosetta energies of −15.4. There was also a substantial mean energy difference between CMA-ES and Genetic of −10.5.

The primary objective of the scaffold matching algorithm is to align a scaffold to hotspot residues. The scaffold is often a peptidomimetic

FIGURE 3 Highlighted examples show covariance matrix adaptation evolution strategy (CMA-ES) outperforms alternative algorithms. Benchmark examples where CMA-ES outperforms Monte Carlo Gradient-based Minimizer (Min), and Genetic algorithms. Row 1: 5-mer peptide from PDB 1NVR. (A) RMSD versus Rosetta Energy scatter plot shows top CMA-ES poses (blue triangles) have lower energy than poses generated with Monte Carlo (orange squares), Minimization (green circle), or Genetic algorithm (pink x) (B) Alignment of lowest energy pose backbones CMA-ES (blue), Monte Carlo (orange), Minimization (green), and Genetic (pink) aligned to native (gray). (C) Alignment of CMA-ES pose to native backbone shows four satisfied hotspot constraints (yellow). Primary hotspot is highlighted in red. Row 2: 10-mer peptide from PDB 1T7R. Same colors as row 1. (D) RMSD versus Rosetta Energy scatter plot shows CMA-ES poses have lower energy and lower RMSD than Monte Carlo, Minimization, and Genetic algorithm. (E) Lowest energy CMA-ES pose aligns very well to native compared with other lowest energy alternative algorithm poses. (F) CMA-ES pose aligned to native satisfies 4 hotspot constraints. Row 3: 9-mer peptide from PDB 2B1Z. Same colors as row 1. (G) RMSD versus Rosetta Energy scatter plot shows nearly all CMA-ES poses have lower energy and lower RMSD than Monte Carlo, Minimization, and Genetic algorithm. (H) Lowest energy CMA-ES pose aligns well with native backbone compared with other lowest energy alternative algorithm poses. (I) CMA-ES pose aligned to native satisfies six hotspot constraints. *Poses with a Rosetta energy score >50 are not shown on scatter plots, number of data points not shown is indicated on plot.

10970134, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/prot.26619 by Helmholtz Zentrum Muenchen Deutsches Forschungszentrum, Wiley Online Library on [07/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

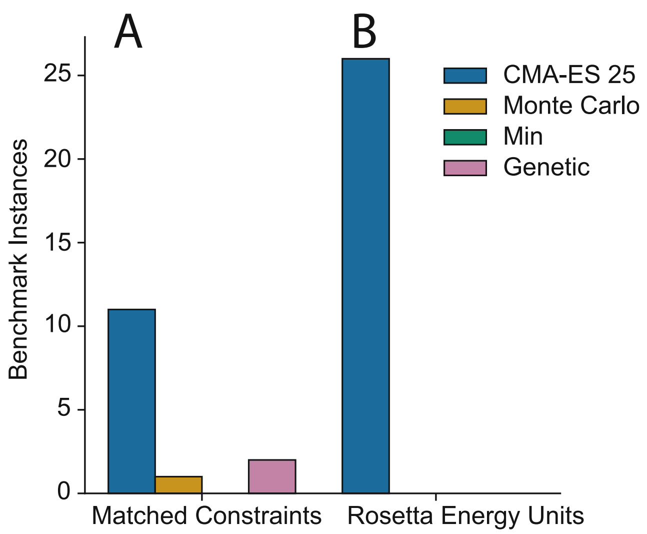

FIGURE 4 Global benchmark analysis showing covariance matrix adaptation evolution strategy (CMA-ES) performance. The bar plot shows the number of benchmark instances where CMA-ES (blue), Monte Carlo (orange), Minimizer (green), or Genetic (pink) algorithms produced the top performing model. (A) Evaluation of top performing models based on the number of matched hotspot residues. CMA-ES matched more hotspot residues than the other algorithms for 11 benchmark instances while Genetic and Monte Carlo produced the top model for only 2 and 1 benchmark instances, respectively. (B) Evaluation of top performing models based on Rosetta energy. CMA-ES models scored lower energies for all 26 Benchmark instances.

which does not have a native pose to compare to. Additionally, to evaluate an optimization algorithm such as CMA-ES, the goal is to find the lowest energy conformation given the provided energy function irrespective of the native conformation. As shown above for individual examples, RMSD to the native; however, still serves as a valuable secondary metric to evaluate when available. Here, we evaluate the CMA-ES algorithm in terms of the lowest energy pose's root mean square deviation (RMSD) to the native peptide. Overall, when evaluating RMSD to the native, we did not identify an algorithm that outperformed all others in all cases (Figure S4M–R). We do note however that the Monte Carlo algorithm shows a slightly better performance than CMA-ES (Figure S4M). One outlier in this distribution is the benchmark instance 1SSH. CMA-ES lowest energy pose has an RMSD to the native peptide of ~15.9 Å while Monte Carlo has an RMSD of ~7.4 Å for a difference of ~8.5 Å. When visually inspecting the 3D structure of the pose in comparison to native for these two final poses (Figure S3, 1SSH), we can see that CMA-ES has sampled much more broadly than Monte Carlo. Interestingly, the total Rosetta energy for the CMA-ES pose is lower than the Monte Carlo pose (~10.7 vs. ~17.6 REU) and additionally the CMA-ES matches more hotspot residues than the Monte Carlo (2 vs. 1). This example shows CMA-ES can find low-energy conformations far from the native that satisfy multiple hotspot residues.

A direct comparison of CMA-ES with the Minimizer algorithm; however, shows a slight majority of CMA-ES models are more similar to the native peptide (Figure S4N). CMA-ES also outperformed the Genetic algorithm, again with a slight majority of models more similar to the native conformation (Figure S4O).

Using these standard performance metrics, we see CMA-ES is a robust optimization algorithm. We see that it outcompetes the standard default Rosetta Minimizer, a customized Monte Carlo algorithm, and the Genetic algorithm using Rosetta energy score as an evaluation metric. It also outcompetes the Minimizer and the Genetic algorithm in terms of RMSD to the native peptide. Finally, CMA-ES matches far more hotspot residues than any of the other algorithms.

2.6 | Comparison of CMA-ES sigma values

Apart from the coordinates of the scaffold's initial conformation, the CMA-ES algorithm requires a sigma value which is the initial standard deviation for the multivariate normal distribution. Although the algorithm automatically updates the sigma value for each degree of freedom after every iteration to find an optimal value, we wanted to evaluate the performance of CMA-ES for different values of sigma. We therefore ran the test workflow using sigma values of 0.3, 10, 25, and 60. We identified the sigma value of 25 generated poses with lower total Rosetta energy in 20 benchmark instances when compared with sigma value of 0.3 and 10, and lower total Rosetta energy in 17 benchmark instances when compared with sigma value of 60 (Table S1).

As one increases the sigma value, the algorithm samples more widely. A noteworthy example of this is the benchmark instance PDBID 1RXZ. In Figure S3 1RXZ, we observe that CMA-ES samples conformational space that is inaccessible to the Monte Carlo, Minimizer, and Genetic algorithms. In Figure S5, we show the low-energy poses from CMA-ES runs with increasing sigma values for 1RXZ. We can see the CMA-ES algorithm first conservatively sampling close to the initial structure with sigma values of 0.3 and 10. When a sigma value of 25 is used we see considerably more sampling with portions of the pose mimicking the native peptide structure. We see in Table S1, the CMA-ES run with sigma values of 0.3 and 10 matched only a single hotspot residue. When the sigma value was increased to 25, CMA-ES was able to identify a conformation that matched 3 hotspot residues suggesting the increased sampling allowed access into conformational space in alternative energy minima.

Further, we observe results run with sigma value of 60 increase conformational sampling, yet results are variable. In particular, the final poses from sigma = 60 runs match hotspots missed by runs with lower values of sigma but also miss hotspots matched by other runs as well. For example, CMA-ES with sigma = 60 produced a low-energy pose with 5 matched hotspots for PDB 1ER8, where in comparison no other algorithm or CMA-ES sigma value matched more than 1 hotspot (the primary hotspot). However, for PDB 2P54 CMA-ES sigma 60 produced a low-energy pose with only 1 matched hotspot, but CMA-ES at the other sigma values matched 7 hotspots. Looking again at PDB 1RXZ (Figure S5), the low-energy pose generated by sigma = 60 produced a conformation that was substantially variable from the initial input pose but did not match any hotspots.

10970134, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/prot.26619 by Helmholtz Zentrum Muenchen Deutsches Forschungszentrum, Wiley Online Library on [07/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

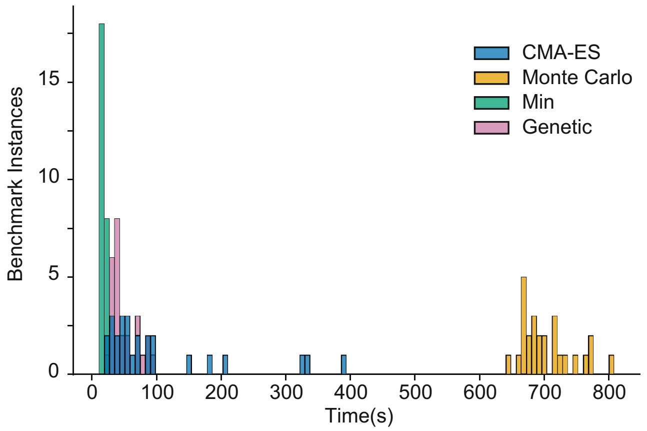

FIGURE 5 Run time analysis demonstrates covariance matrix adaptation evolution strategy (CMA-ES) is competitive with Minimizer algorithm. Histograms of runtimes for benchmark instances run using each optimization algorithm: CMA-ES (blue), Monte Carlo (orange), Minimizer (green), and Genetic (pink). The Minimizer algorithm had a median runtime of 15 s, while CMA-ES and Genetic had slightly higher median runtimes of 57 and 42 s, respectively. Monte Carlo had a substantially slower median runtime of 695 s.

We decided a sigma value of 25 was optimal for our purposes, as it produced poses with more matched hotspots and lower energy values on average.

2.7 | Time comparison of algorithms

With most optimization algorithms, there is a tradeoff between speed and sampling. We next evaluate the algorithms in terms of their run time to evaluate their efficiency to better understand this tradeoff. As can be seen in Figure 5, the Rosetta Minimizer has the fastest runtime with a median runtime of 15 s among all benchmark instances. The CMA-ES and Genetic algorithms have comparable median runtimes of 57 and 42 s while the Monte Carlo algorithm has an order of magnitude longer median time of 695 s. We therefore conclude based on the time analysis that the CMA-ES algorithm efficiently samples the energy landscape and effectively identifies low-energy conformations appropriately balancing the speed/sampling tradeoff.

2.8 | Matching of peptidomimetic scaffold to SARS-CoV-2 mPro substrate residues

Now that we have established CMA-ES as an efficient algorithm to match scaffolds on to hotspot residues, we next demonstrate the ability of the full Scaffold Matcher algorithm in conjunction with CMA-ES to match peptidomimetic scaffolds onto the substrate residues of the SARS-CoV-2 mPro. The SARS-CoV-2 mPro is a 3C-like protease, which process nonstructural proteins that function as the replication and transcriptional machinery during the viral life cycle.[22] MPro is therefore a primary target for antiviral drug therapeutics. The peptide substrate fits into the mPro binding groove with several substrate residues making high-affinity contacts including a buried Leucine two residues N-terminal of the cleavage site (position P2) and a Valine (position P3; Figure 6A). We therefore used the native Leucine rotamer as the primary hotspot residue and the native Valine rotamer as the ancillary hotspot residue. We then used the Scaffold Matcher algorithm with CMA-ES to dock an oligooxopiperzine scaffold onto these two hotspot residues. The top matching pose is shown in Figure 6B and aligns well within mPro's substrate binding groove suggesting it would perform as an inhibitor blocking an interaction between mPro and its substrate peptide. We can also see good alignment of the oligooxopiperazine with the hotspot residues shown in complex with mPro (Figure 6C) as well as rotated with the complex hidden (Figure 6D). We finally compare the performance of CMA-ES versus the gradient descent minimizer on identifying low-energy poses that match multiple hotspots. In Figure 6E, we show CMA-ES identifies low-energy poses that better satisfy the hotspot constraints (i.e., Hotspot Constraint Score). This proof-of-concept demonstration shows how the CMA-ES algorithm can be used to dock peptidomimetics onto the hotspot residues of therapeutically relevant drug targets and how the resulting docked molecule could be used for future lead development.

3 | MATERIALS AND METHODS

3.1 | Scaffold Matcher implementation

The Scaffold Matcher algorithm is implemented in the Rosetta macromolecular modeling software suite. A full pseudocode description of the algorithm can be found in the Supplementary Methods (File S1). Briefly, the algorithm is given a molecular scaffold (e.g., oligooxopiperazine, peptide) and a set of hotspot residues coordinates. In addition, the algorithm is passed in the indices of the primary hotspot residue and the anchor position of the molecular scaffold. The algorithm first aligns the anchor position of the scaffold to the primary hotspot coordinates. Constraints are then created for each additional hotspot residue (i.e., ancillary) to all other scaffold positions. The constraints are formulated such that if a position on the molecular scaffold aligns with the hotspot coordinates, the total energy of the system is lowered (additional details are described in Supplementary Methods, File S1). The CMA-ES optimization algorithm (see below) is then used to traverse the energy landscape to identify a molecular conformation with low energy and that satisfies the hotspot constraints.

Below is an example commandline to match an idealized 8mer peptide (i.e., ALA8.pdb) to hotspot residues in the peptide in PDBID 1N7F. To simplify the evaluation framework, we utilize a blank target peptide (i.e., AAAAAA_blank.pdb) to represent the binding partner which is distant in space from the hotspots and does not interact with the molecular scaffold nor change conformation. The anchor position (i.e., 10) is relative to all positions in the system including the target peptide. The primary hotspot (i.e., primary_hs_stub_lib) and ancillary hotspots (ancillary_hs_stub_libs) are pdb formatted files with coordinates the hotspot residues. Here, we evaluate only the native coordinates. The command generates 100 independent structures (i.e., nstruct). All input files and example output files can be found at

10970134, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/prot.26619 by Helmholtz Zentrum Muenchen Deutsches Forschungszentrum, Wiley Online Library on [07/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

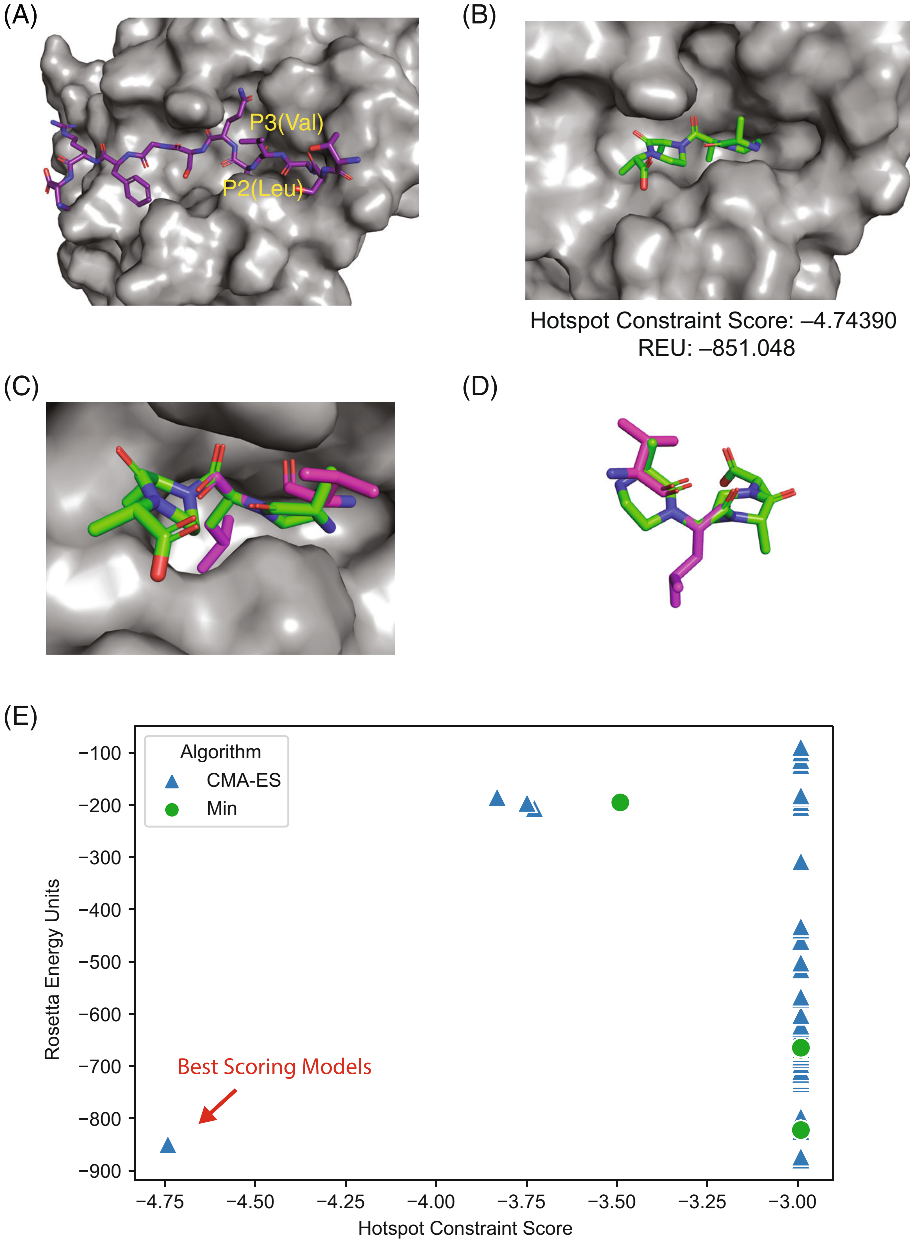

FIGURE 6 Mimicking the interface of SARS-CoV-2 mPro. (A) SARS-CoV-2 mPro (6Y2E) with substrate peptide aligned from 2Q6G. (B) Oligooxopiperazine scaffold docked into substrate pocket of SARS-CoV-2 mPro using covariance matrix adaptation evolution strategy (CMA-ES). (C) Docked oligooxopiperazine scaffold shown with matched hotspots VAL4 and LEU5. (D) Alternative view of oligooxopiperizine scaffold aligned to hotspots without mPro shown.
(E) Scatterplot of Hotspot Constraint Score versus Rosetta Energy Units shows CMA-ES identifies lower energy-matched scaffold conformations with better matched hotspots (blue triangles) in comparison to gradient-based minimizer (green circles). Red arrow points to models of best scoring models (lowest energy and top hotspot matches) where only CMA-ES scaffold conformations but no Minimizer conformations are represented.

https://doi.org/10.5281/zenodo.8422476 in ScaffoldMatcherBenchmarkFiles_20230410.zip.

```
~/bin/scaffold_matcher.default.linuxgccrelease -s
./ALA8.pdb -hstarget ./AAAAAA_blank.pdb -nstruct
100 -primary_hs_stub_lib ./1N7F_BD_D90_natOnly_lib.
pdb -ancillary_hs_stub_libs ./1N7F_BD_D87_lib.pdb
./1N7F_BD_D88_lib.pdb ./1N7F_BD_D89_lib.pdb
./1N7F_BD_D91_lib.pdb ./1N7F_BD_D92_lib.pdb
./1N7F_BD_D93_lib.pdb ./1N7F_BD_D94_lib.pdb
-hs_repack_only false -out:file:scorefile
scaffold_matcher.sc -run:
multiple_processes_writing_to_one_directory
-fixed_anchor 10 -use_cmaes true -use_minimizer false
-optimization:cmaes_rgsigma 25
```

3.2 | CMA-ES implementation in Rosetta

We based our Rosetta CMA-ES optimization implementation on the core c-cmaes library code from[11] (https://github.com/CMA-ES/c-cmaes). The CMA-ES optimization algorithm is first initialized with a vector of degrees of freedom which takes the form of initial means of the multivariate normal distribution. Similar to other minimization algorithms implemented in Rosetta, degrees of freedom represent torsion angles along a peptide backbone (i.e., phi, psi, and omega), side-chain torsion angles (i.e., chi), and rotational/translational degrees of freedom. A vector of sigma values is also initialized representing the starting standard deviation of each degree of freedom in the multivariate normal distribution. We use a sigma value of 25 for the starting standard deviation unless otherwise noted.

10970134, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/prot.26619 by Helmholtz Zentrum Muenchen Deutsches Forschungszentrum, Wiley Online Library on [07/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Once initialized, the algorithm then obtains a population of samples from the multivariate normal distribution of size lambda. We use the recommended lambda value of 4 + int(3*log(*N*)), where N is the number of degrees of freedom. Each sample is evaluated using the selected Rosetta energy score function (ref2015 for this application). The samples are then ranked based on their evaluated energies and the top mu samples are then used to select the new mean and update the covariance matrix of the next iteration's multivariate normal distribution. Mu is set to ½ of lambda. The algorithm iterates until either a selected tolerance parameter of 0.001 is reached or a max number of iterations of 1000 is reached.

A full mathematical description of the CMA-ES algorithm is described in the Supplementary Methods (File S1).

3.3 | Monte Carlo algorithm

We used the Monte Carlo algorithm implementation in Rosetta which is a modification of the protocol used in Renfrew et al.[23] Briefly, the protocol iterates through a perturbation cycle 50 times, randomly making small and shear moves of backbone phi and psi angles of max 2 degrees. All phi and psi angles along the peptide backbone are subject to perturbation. Monte Carlo temperature for perturbation moves is set at 0.8 while each iteration temperature value is set at 0.5. In contrast to the original algorithm in Renfrew et al.,[23] docking (i.e., rotation and translation) is turned off to keep the scaffold fixed on the primary hotspot residue. Also, the design step is turned off as well as no final design minimization step. We generated 100 models.

3.4 | Minimizer algorithm

We used the default gradient-based minimizer in Rosetta which optimizes a set of degrees of freedom for a specified score function. The default minimizer is an implementation of the L-BFGS algorithm (i.e., lbfgs_armijo_nonmonotone).[24] We allowed only backbone torsion angles to move. We set the tolerance parameter to 0.001.

3.5 | Genetic algorithm

We used the GA_Minimizer in Rosetta which implements a genetic algorithm for comparison to CMA-ES. The GA_Minimizer, similar to other minimization algorithms, optimizes a set of degrees of freedom for a specified score function. First, the algorithm creates a population of size 20 by adding random noise to the initial degrees of freedom values and the samples are ranked according to their score. Next, mutation and crossover steps are performed to diversify the samples. Mutations and crossovers both performed with 0.5 probability. The algorithm repeats until the tolerance score threshold of 0.001 is reached.

3.6 | Model tri-peptide CMA-ES minimization

We used pymol to construct an Ala-His-Ala peptide. We used the commandline below to minimize the model tri-peptide.

```
./Rosetta/main/source/bin/minimize.default.
linuxgccrelease -database ./Rosetta/main/database/
-s ./AHA.pdb -run:min_type cmaes
```

3.7 | Benchmark workflow

We downloaded structures of peptide–protein interaction pairs found in the FlexPepDock benchmark (https://doi.org/10.1371/journal.pone.0018934.s002).[21] Each structure was "cleaned" using the pdb_clean.py script provided in the Rosetta tools directory. Structures were then "relaxed" using Rosetta's FastRelax protocol with the "-relax:constrain_relax_to_start_coords" flag. 50 decoys were produced and the lowest scoring model was selected. For each benchmark instance, the peptide was extracted from the target and backbone atoms removed, resulting in a set of disembodied sidechains. Glycine residues were ignored. Each disembodied sidechain was placed in a separate pdb formatted stub constraint library file. Idealized peptides of the same size and nearest secondary structure as the target peptide were generated using PyMOL.[25] A primary hotspot residue with its corresponding peptide anchor position was randomly selected for each peptide and all other disembodied side chains are considered ancillary hotspot residues. The idealized backbone is aligned to the primary hotspot residue. Hotspot constraints are created for the primary hotspot and all ancillary hotspot residues. One hundred decoy samples were run for each peptide and each algorithm (i.e., Monte Carlo, Minimizer, and CMA-ES). The Rosetta score function ref2015 was used for all energy calculations. The root mean squared deviation (RMSD) was calculated between a Rosetta-generated peptide conformation and the native peptide conformation found in the PDB for the respective benchmark instance. Only alpha carbon atoms were considered in the RMSD calculation.

3.8 | Oligooxopiperazine scaffold matched to SARS-CoV-2 mPro peptide substrate hotspots

We modeled the structure of SARS-CoV-2 mPro using the x-ray crystal structure from PDB ID 6Y2E.[26] The substrate peptide from the SARS-CoV mPro (PDB ID 2Q6G)[27] was modeled into the binding pocket of the SARS-CoV-2 mPro by structural superposition of the conserved proteases. Hotspot residue pdb files for Leucine and Valine residues were generated manually by extracting the corresponding substrate peptide residue's *xyz* coordinates. A 4-alanine oligooxopiperazine scaffold was used as input into the Scaffold Matcher algorithm to align onto the two hotspot residues, Leucine and Valine. Leucine was designated the primary hotspot and Valine was the ancillary hotspot. The Scaffold Matcher algorithm was run with the CMA-

10970134, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/prot.26619 by Helmholtz Zentrum Muenchen Deutsches Forschungszentrum, Wiley Online Library on [07/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ES optimizer (sigma = 10) as well as Rosetta's default gradient-based minimizer for comparison purposes, both generating 100 poses each. Input files can be found at https://doi.org/10.5281/zenodo.8422476 in ScaffoldMatcherBenchmarkFiles_20230410.zip.

3.9 | Commandline for scaffold matcher with CMA-ES for SARS-CoV-2 run

```
./bin/scaffold_matcher.default.linuxgccrelease -s
./oop_dimer_LLLL_min.pdb -hstarget ./COVI_A.pdb
-nstruct 100 -primary_hs_stub_lib
./COVI_AC_C311_native_lib.pdb
-ancillary_hs_stub_libs ./COVI_AC_C310_native_lib.
pdb -hs_repack_only false -out:file:scorefile
scaffold_matcher.sc -run:
multiple_processes_writing_to_one_directory
-optimization:cmaes_rgsigma 10
```

All associated code is freely accessible to academic users via the ROSETTACOMMONS website (http://www.rosettacommons.org).

4 | DISCUSSION

Herein, we describe the implementation of the Scaffold Matcher algorithm which aligns molecular scaffolds onto hotspot residues from a protein interaction interface. To optimize the multiple degrees of freedom within the molecular scaffold during the alignment procedure, we implemented a state-of-the-art evolutionary optimization algorithm, CMA-ES into the Rosetta macromolecular modeling framework.

We first demonstrate the CMA-ES algorithm's utility by minimizing the energy of a model tri-peptide. We show that CMA-ES can escape local energy minima while lowering the overall energy. In addition, we observe that CMA-ES efficiently samples degrees of freedom to not waste computation on unnecessary sampling.

We also analyze CMA-ES performance within the Scaffold Matcher algorithm on the task of aligning peptide scaffolds to hotspot residues on a benchmark of 26 peptides. We demonstrate CMA-ES outperforms gradient-based minimization, a Genetic algorithm, and a Monte Carlo algorithm on several evaluation metrics including total number of matched hotspot residues and overall decrease in energy. Interestingly, we observed the Monte Carlo algorithm slightly outperform CMA-ES with respect to RMSD to native (Figure S4G). This result may be due to the Monte Carlo sampling procedure using Rosetta small and shear moves which sample from the Ramachandran distribution of phi and psi angles. Sampling from the Ramachandran distribution is likely to ensure Monte Carlo poses are in a more peptide-like conformation. CMA-ES on the other hand samples from a multivariate normal distribution across all degrees of freedom and therefore has no prior information on peptide geometry other than when a pose is evaluated by the Rosetta energy function. Minor inaccuracies in the Rosetta energy function may be the cause for the discrepancy between peptide geometry and low-energy values. Further investigation is needed to determine why CMA-ES underperforms in the RMSD metric in our analysis.

Finally, we evaluated all four algorithms in terms of their runtime. While gradient-based minimization performed the fastest as expected, CMA-ES was highly competitive on most of the test cases (Figure 5). Also as expected, the Monte Carlo algorithm performed slowest. Energy optimization is often a tradeoff between speed and finding lower energy minima. CMA-ES balances this tradeoff by consistently finding lower energy minima than the gradient-based minimization algorithm, the Genetic algorithm, and Monte Carlo algorithm as well as having reasonable run times.

Unlike some other evolutionary algorithms, CMA-ES does not require many parameters to be specified at runtime. A user only provides an initial starting conformation and initial sigma value. The sigma value is automatically updated after each iteration to determine optimal values, so the algorithm is robust to initial values as we describe above. A user may want to guide CMA-ES toward increased/decreased sampling and therefore can set sigma higher or lower than our suggested value of 25. A population value lambda can be modified but the default value which is a function of the number of degrees of freedom has been shown to work well in most cases. Additional strategies are being explored to increase the population size after restart which has shown additional performance improvements.[28] Moreover, hybrid strategies that combine gradient-based optimization and evolution strategies also show great promise in boosting performance.[29] Overall, however, the algorithm in its current state requires minimal to no tuning from the user. This points to its high-value in many future modeling applications including ligand docking,[30] peptidomimetic modeling[31] and design,[7] and cyclic peptide design.[32–34]

Finally, we demonstrate our Scaffold Matcher algorithm on mimicking the hotspot residues from the peptide substrate of the SARS-CoV-2 mPro. The mPro protein is essential for the lifecycle of the SARS-CoV-2 virus and developing an inhibitor has been an active area of interest since early in the pandemic.[35] Our approach allows for the identification of molecular scaffolds that hold the beneficial properties of peptidomimetics (e.g., proteolytic stability, synthetic tractability), and mimic the hotspot residues at protein interaction interfaces. We anticipate the Scaffold Matcher algorithm along with CMA-ES will greatly speed our ability to identify molecular scaffolds for inhibitors and ultimately develop leads for therapeutic targets.

AUTHOR CONTRIBUTIONS

Erin R. Claussen: Software; writing – review and editing; visualization; validation; methodology; writing – original draft; investigation; formal analysis; data curation. **P. Douglas Renfrew:** Conceptualization; investigation; methodology; validation; visualization; writing – review and editing; formal analysis; supervision; resources; data curation; software. **Christian L. Müller:** Conceptualization; investigation; writing – review and editing; visualization; validation; methodology; software; formal analysis; data curation. **Kevin Drew:** Conceptualization; investigation; funding acquisition; writing – original draft; methodology; validation; visualization; writing – review and editing;

10970134, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/prot.26619 by Helmholtz Zentrum Muenchen Deutsches Forschungszentrum, Wiley Online Library on [07/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

software; formal analysis; project administration; data curation; supervision; resources.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health (R00 HD092613 and L40 HD096554 to Kevin Drew) and RosettaCommons (RC22024 to Kevin Drew.). The authors would also like to thank José Villegas and Parisa Hosseinzadeh for helpful discussions, Matteo Ferla for the initial mPro pdb model, and Sergey Lyskov for his implementation of the Genetic algorithm in Rosetta.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/prot.26619.

DATA AVAILABILITY STATEMENT

The data used for this work can be found at https://zenodo.org/record/8422476 which includes input files and top models. The Rosetta Software Code can be found at https://www.rosettacommons.org.

ORCID

Kevin Drew https://orcid.org/0000-0002-1260-4413

REFERENCES

1. Goh K-I, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL. The human disease network. *Proc Natl Acad Sci U S A*. 2007;104:8685-8690.
2. Zinzalla G, Thurston DE. Targeting protein–protein interactions for therapeutic intervention: a challenge for the future. *Future Med Chem*. 2009;1:65-93.
3. Pelay-Gimeno M, Glas A, Koch O, Grossmann TN. Structure-based design of inhibitors of protein–protein interactions: mimicking peptide binding epitopes. *Angew Chem Int Ed*. 2015;54:8896-8927.
4. Thanos CD, DeLano WL, Wells JA. Hot-spot mimicry of a cytokine receptor by a small molecule. *Proc Natl Acad Sci U S A*. 2006;103:15422-15427.
5. Leman JK, Weitzner BD, Lewis SM, et al. Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat Methods*. 2020;17:665-680.
6. Drew K, Renfrew PD, Craven TW, et al. Adding diverse noncanonical backbones to Rosetta: enabling peptidomimetic design. *PloS One*. 2013;8:e67051.
7. Lao BB, Drew K, Guarracino DA, et al. Rational design of topographical helix mimics as potent inhibitors of protein–protein interactions. *J Am Chem Soc*. 2014;136:7877-7888.
8. Leaver-Fay A, Tyka M, Lewis SM, et al. Rosetta3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol*. 2011;487:545-574.
9. Cheng RP, Gellman SH, DeGrado WF. β-Peptides: from structure to function. *Chem Rev*. 2001;101:3219-3232.
10. Hansen N, Ostermeier A. Completely derandomized self-adaptation in evolution strategies. *Evol Comput*. 2001;9:159-195.
11. Hansen N. The CMA evolution strategy: a comparing review. 28.
12. Szynkiewicz P. A comparative study of PSO and CMA-ES algorithms on black-box optimization benchmarks. *J Telecommun Inf Technol*. 2019;4:5-17.
13. Rakhshani H, Idoumghar L, Lepagnot J, Brévilliers M. Speed up differential evolution for computationally expensive protein structure prediction problems. *Swarm Evol Comput*. 2019;50:100493.
14. Winter R, Montanari F, Steffen A, Briem H, Noé F, Clevert DA. Efficient multi-objective molecular optimization in a continuous latent space. *Chem Sci*. 2019;10:8016-8024.
15. Maximova T, Moffatt R, Ma B, Nussinov R, Shehu A. Principles and overview of sampling methods for modeling macromolecular structure and dynamics. *PLoS Comput Biol*. 2016;12:e1004619.
16. Desjarlais JR, Handel TM. De novo design of the hydrophobic cores of proteins. *Protein Sci Publ Protein Soc*. 1995;4:2006-2018.
17. Raha K, Wollacott AM, Italia MJ, Desjarlais JR. Prediction of amino acid sequence from structure. *Protein Sci Publ Protein Soc*. 2000;9:1106-1119.
18. Clausen R, Sapin E, De Jong KA, Shehu A. Evolution strategies for exploring protein energy landscapes. Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation, 217–224 (Association for Computing Machinery) 2015. doi:10.1145/2739480.2754692
19. Müller CL, Sbalzarini IF. Energy landscapes of atomic clusters as black box optimization benchmarks. *Evol Comput*. 2012;20:543-573.
20. Fleishman SJ, Whitehead TA, Ekiert DC, et al. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science*. 2011;332:816-821.
21. Raveh B, London N, Zimmerman L, Schueler-Furman O. Rosetta FlexPepDock ab-initio: simultaneous folding, docking and refinement of peptides onto their receptors. *PLoS One*. 2011;6:e18934.
22. Douangamath A, Fearon D, Gehrtz P, et al. Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nat Commun*. 2020;11:5047.
23. Renfrew PD, Choi EJ, Bonneau R, Kuhlman B. Incorporation of noncanonical amino acids into Rosetta and use in computational protein-peptide Interface design. *PLoS One*. 2012;7:e32637.
24. Liu DC, Nocedal J. On the limited memory BFGS method for large scale optimization. *Math Program*. 1989;45:503-528.
25. Schrödinger, LLC. *The PyMOL Molecular Graphics System, Version 2.0*. Schrödinger, LLC. 2015.
26. Zhang L, Lin D, Sun X, et al. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science*. 2020;368:409-412.
27. Xue X, Yu H, Yang H, et al. Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. *J Virol*. 2008;82:2515-2527.
28. Auger A, Hansen N. A restart CMA evolution strategy with increasing population size. 2005 IEEE Congress on Evolutionary Computation. (vol. 2, 1769–1776).
29. Maheswaranathan N, Metz L, Tucker G, Choi D, Sohl-Dickstein J. Guided evolutionary strategies: augmenting random search with surrogate gradients. Proceedings of the 36th International Conference on Machine Learning, 4264–4273 (PMLR, 2019) 2019.
30. Fu DY, Meiler J. RosettaLigandEnsemble: a small-molecule ensemble-driven docking approach. *ACS Omega*. 2018;3:3655-3664.
31. Butterfoss GL, Yoo B, Jaworski JN, et al. De novo structure prediction and experimental characterization of folded peptoid oligomers. *Proc Natl Acad Sci U S A*. 2012;109:14320-14325.
32. Hosseinzadeh P, Watson PR, Craven TW, et al. Anchor extension: a structure-guided approach to design cyclic peptides targeting enzyme active sites. *Nat Commun*. 2021;12:3384.
33. Rettie SA, Campbell KV, Bera AK, et al. Cyclic peptide structure prediction and design using AlphaFold. 2023. 2023.02.25.529956 Preprint at:. doi:10.1101/2023.02.25.529956
34. Soler MA, Rodriguez A, Russo A, et al. Computational design of cyclic peptides for the customized oriented immobilization of globular proteins. *Phys Chem Chem Phys*. 2017;19:2740-2748.

10970134, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/prot.26619 by Helmholtz Zentrum Muenchen Deutsches Forschungszentrum, Wiley Online Library on [07/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

35. Zaidman D, Gehrtz P, Filep M, et al. An automatic pipeline for the design of irreversible derivatives identifies a potent SARS-CoV-2 Mpro inhibitor. *Cell Chem Biol*. 2021;28:1795-1806.e5.

SUPPORTING INFORMATION

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

How to cite this article: Claussen ER, Renfrew PD, Müller CL, Drew K. Scaffold Matcher: A CMA-ES based algorithm for identifying hotspot aligned peptidomimetic scaffolds. *Proteins*. 2024;92(3):343-355. doi:10.1002/prot.26619

10970134, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/prot.26619 by Helmholtz Zentrum Muenchen Deutsches Forschungszentrum, Wiley Online Library on [07/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License