Bioinformatics, 2024, **40(8)**, btae492
https://doi.org/10.1093/bioinformatics/btae492
Advance Access Publication Date: 6 August 2024
Applications Note

Structural bioinformatics

Locuaz: an *in silico* platform for protein binders optimization

German P. Barletta [1,2,*], Rika Tandiana [1], Miguel Soler [1,3], Sara Fortuna [1,4,†,*], Walter Rocchia [1,†,*]

[1]CONCEPT, Istituto Italiano di Tecnologia, Via Enrico Melen, 83 Genova Liguria 16152, Italy
[2]The Abdus Salam International Centre for Theoretical Physics (ICTP), Str. Costiera, 11, Trieste, Friuli-Venezia Giulia, 34151, Italy
[3]Dipartimento di Scienze Matematiche, Informatiche e Fisiche (DMIF), University of Udine, Via delle Scienze, 206, Udine, Friuli-Venezia Giulia, 33100, Italy
[4]Cresset, New Cambridge House, Litlington, Royston, SG8-0SS, United Kingdom

*Corresponding authors. CONCEPT, Istituto Italiano di Tecnologia, Via Enrico Melen, 83 Genova Liguria 16152, Italy. Emails: walter.rocchia@iit.it (W.R.), sara.fortuna@cresset-group.com (S.F.), pbarletta@gmail.com (G.P.B.)

†Contributed equally to this work

Associate Editor: Inanc Birol

Abstract

Motivation: Engineering high-affinity binders targeting specific antigenic determinants remains a challenging and often daunting task, requiring extensive experimental screening. Computational methods have the potential to accelerate this process, reducing costs and time, but only if they demonstrate broad applicability and efficiency in exploring mutations, evaluating affinity, and pruning unproductive mutation paths.

Results: In response to these challenges, we introduce a new computational platform for optimizing protein binders towards their targets. The platform is organized as a series of modules, performing mutation selection and application, molecular dynamics simulations to sample conformations around interaction poses, and mutation prioritization using suitable scoring functions. Notably, the platform supports parallel exploration of different mutation streams, enabling *in silico* high-throughput screening on High Performance Computing (HPC) systems. Furthermore, the platform is highly customizable, allowing users to implement their own protocols.

Availability and implementation: The source code is available at https://github.com/pgbarletta/locuaz and documentation is at https://locuaz.readthedocs.io/. The data underlying this article are available at https://github.com/pgbarletta/suppl_info_locuaz

1 Introduction

The antibody (Ab) discovery process, a critical aspect of biotherapeutics development, relies on the identification of one or more starting candidates, for instance via phage display, yeast display, and hybridoma technology. The affinity maturation follows through several steps of either site-directed mutagenesis, directed evolution, or deep mutational scanning which allow assessing the impact of specific mutations on Ab function and stability (Kennedy *et al.* 2018). The goal is to obtain an Ab with optimal stability, high affinity and specificity toward the target. *In silico* approaches can be extremely useful in this context, especially when structural data of the target are available. They include knowledge-based methods, trained on sequence and structure databases, physics-based methods, and hybrid approaches, all aiming at mimicking the details of protein-protein interactions at a lower cost and a shorter time (Sormanni *et al.* 2018).

Many popular empirical methods belong to the Rosetta suite(Sivasubramanian *et al.* 2009), which has been used extensively, also in some case to optimize CDR3 sequences, producing Abs that were readily expressed and validated (Moriyama *et al.* 2023). For instance, Rosetta RAbD, an iterative protocol generating redesigned mutants and samples conformations via a Monte Carlo scheme using a specific energy definition, led to the discovery of an Ab with a Kd ∼20 nM (Adolf-Bryfogle *et al.* 2018). This result is comparable to those of experimental methods, which very rarely deliver Abs with pM binding affinity.

OSPREY (Hallen *et al.* 2018) is another well-established software which has proven capable of achieving a 100-fold increase in Ab affinity with just a single point mutation (Holt *et al.* 2023).

Physics-based approaches often rely on molecular dynamics (MD) simulations for conformational sampling. Among them, a protocol based on MD coupled with a Metropolis Montecarlo scheme using energies calculated via MMPBSA to accept sampled conformations led to an Ab with Kd ∼0.5 nM (Buratto *et al.* 2022). MD+FoldX is another method where the interaction energies are calculated exclusively with FoldX. It has been applied to Abs against the SARS-Cov-2 receptor binding domain identifying both mutations that could enhance Ab affinity and positions that could potentially host Ab escape mutations. (Barnes *et al.* 2022) In other approaches, the conformations generated via MD were selected with an acceptance criterion based on a consensus

Received: 30 March 2024; **Revised:** 28 July 2024; **Editorial Decision:** 31 July 2024; **Accepted:** 3 August 2024

score among multiple scoring functions giving raise to an evolutive optimization protocol for nanobodies and peptides, reaching affinities comparable to those of experimental techniques (Soler *et al.* 2019, Ochoa *et al.* 2021). A later implementation exploiting replica-exchange MD allowed the *de novo* optimisation of Ab fragments (Soler *et al.* 2023).

Among the most advanced examples of knowledge-based methods, certainly those based on machine learning deserve to be mentioned, such as generative methods. In a recent effort, 1 million Abs against the HER 2 growth factor receptor have been generated in a zero-shot fashion, without any training sample binding HER 2 or one of his homologs. From these designs, 421 binders were experimentally validated and three displayed stronger binding than trastuzumab, the Ab binder licensed as a drug (Shanehsazzadeh *et al.* 2023). Geometric deep learning is another interesting approach which has been experimentally proven to generate optimized CDR sequences of existing Abs (Shan *et al.* 2022). Along the same lines, a deep learning based method trained on 10^4 Ab variants of the antiHer2 Ab Trastuzumab, was used to pick among 10^8 mutants achieving affinities of Kd 0.1–10 nM, thus comparable to that of the original Ab they were derived from (Mason *et al.* 2021), Language models too have been used to optimise Abs with Kd <1nM (Hie *et al.* 2023).

Overall, deep-learning based methods are very promising but need a very large training set of experimental data, which might not always be available. These results highlight that *in silico* methods could be key to avoid massive experimental costs. Moreover, the variety of available protocols and of their intermediate steps allows to tailor each phase of the process, such as the mutation strategy or the affinity estimation criteria.

We propose here *locuaz*: A flexible, python-based platform whose primary goal is to optimize the binding affinity of an Ab (fragment) towards its target. In an evolutionary framework, *locuaz* mutates the candidate binder, samples conformations via MD and scores the affinity of the new complex. Each of these steps can be customized by the user. Importantly, the entire mutation pipeline allows the concurrent generation of different mutation lineages in parallel. The platform can be containerized and run through different job scheduling systems, such as SLURM or PBS. It was constructed using a modular approach both by selecting features from existing Ab evolutionary protocols and by developing new ones into a unique platform, in a way to facilitate updates, customizations and extensions.

2 Materials and methods

As shown in Fig. 1A, the workflow of *Locuaz* is organized in functional units called Blocks, namely the **Mutation Chooser**, which determines which mutation has to be applied, the **Mutator**, which actually applies the mutation to the structure, the **Sampler**, which performs MD simulation to relax the system, the **Complex Scorer**, which evaluates the effect of the current mutation on the binding affinity, and finally the **Pruner**, which decides whether a lineage should be stopped because it is not promising or not. These units were envisioned to support the interchangeable use of built-in tools or of third-party external programs for each task. For example, *Locuaz* currently supports 9 different **Scorers**, 2 possibilities for the input topologies (Amber and GROMACS), GROMACS as MD engine used by the **Sampler**, 3 different **Pruners** and a highly-configurable **Mutation Chooser**. Different flavors of each block can be selected and combined at will, and new ones can be added, leading to a remarkable protocol flexibility.

The optimization process starts from a putative binder/target complex and requires as input the selection of the portion of the binder to be optimized. The workflow then envisions the identification of multiple mutations by the **Mutation Chooser** block. The mutation sites can be one or many, can be either completely random, or guided by biological knowledge, or by physically-based methods, such as MMPB(GB)SA, which are used to recognize which residues are contributing less to the binding affinity (Valdés-Tresanco *et al.* 2021). The choice of the new amino acid is also configurable. The user can choose from different amino acid probability schemes, or assign each amino acid a custom probability. The **Mutation Chooser** also allows custom grouping of amino acids in order to force the choice of the amino acid to be within a certain category of amino acid like" polar" or"

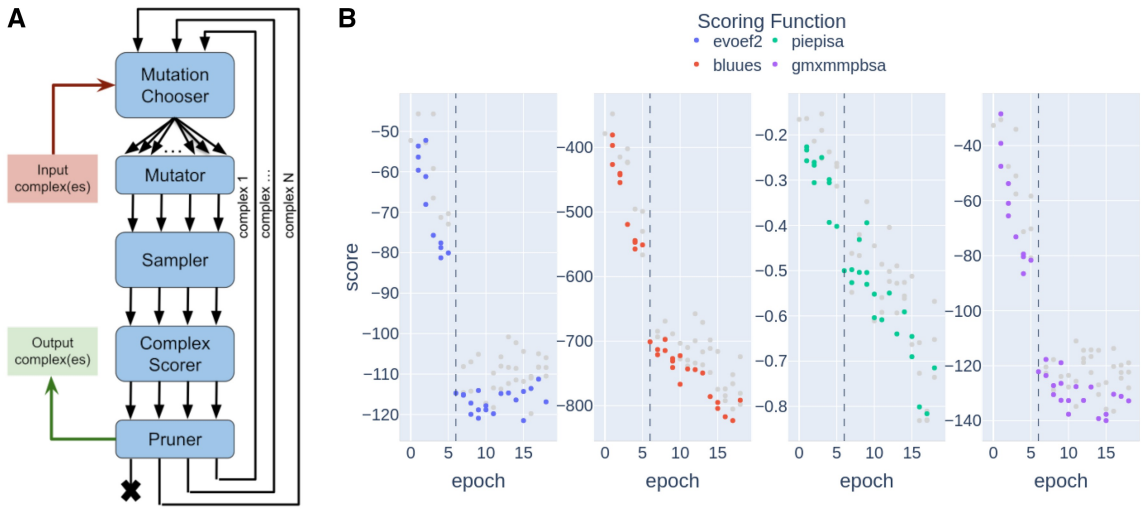

Figure 1. (A) User configurable Blocks of the protocol. (B) Protocol output along the optimization of an anti-p53 Ab fragment: Average scores for all complexes generated along the protocol. Gray markers correspond to complexes that were not selected for the next epoch. Dotted line marks the first epoch of the unrestrained optimization, showing that the lifting of restrains allowed the interface to find a lower energy interaction.

aliphatic", etc More information is available on the documentation.

Mutations are performed by any of the currently available **Mutators** (Soler *et al.* 2018, Tandiana *et al.* 2024).

After a mutation is applied, the **Sampler** performs MD simulations using the GROMACS engine (Abraham *et al.* 2015) by adopting either GROMACS topologies, or those from Amber's Tleap (Salomon-Ferrer *et al.* 2013), allowing the user to include non-standard residues, organic small molecules and ions. To maximize the throughput, all available GPU and CPU resources are pooled and the simulations are distributed among parallel branches (as many as the user requested) and are queued up for concurrent execution following a Producer/Consumer strategy where, in case the number of requests cannot be fulfilled by the available resources, a "first come, first served" criterion is adopted.

Then, the sampled complex configurations are scored and their scores averaged. The score calculations are also dispatched to the available resources, as previously described, to optimize the overall efficiency.

Finally, the user-selected **Pruner** compares the scores of the mutated binders against the original ones and selects the subset of mutated binders that are deemed to improve affinity. If a single scoring function is used, one possibility is to adopt a Monte Carlo-based Pruner (Soler *et al.* 2017), giving a chance also to mutations that are predicted to lead to slightly lower affinity. If, conversely, multiple scoring functions are employed, a consensus criterion **Pruner** can be used to decide which mutation **branch** is to be continued(Soler *et al.* 2019). In case no mutation is retained, the original binders are reused to generate a new set of mutants.

Each cycle of this protocol is called an **epoch**, while simultaneous mutations give rise to different **branches**. A typical optimization process involves many epochs, as depicted on Fig. 1B. At each epoch a user-defined number of branches are generated, effectively forming a Directed Acyclic Graph (DAG). The width of this DAG, i.e. the number of active branches, is managed according to the available computational resources. In Locuaz we included two methods: **variable-width DAG**, suitable when considerable computational resources are available, and **constant-width DAG**, in which the creation of new branches is limited by a previously selected constant width.

3 Application to the TWIST1 system

TWIST1 is a transcription factor which promotes the MDM2-mediated degradation of p53, one of the main tumor suppressors, by interacting with its binding site on p53, known as "Twist-box." A llama nanobody (VHH) shown to interfere with TWIST1: P 53 interaction by binding to p53 with moderate binding affinity was modelled by homology and then docked onto p53 (D'Agostino *et al.* 2022).

In order to improve the p53: VHH affinity, the complex was submitted to 5 epochs of optimization using positional restraints between the binder and the target. Then, 250 ns of NPT were carried out to confirm the improved stability. The most representative structure was chosen after clustering, and used to start another 13 epochs of unrestrained optimization for a total of 18 epochs. The optimisation process was monitored by plotting the 4 selected scoring functions along the epochs (Fig. 1B). In this particular example the **Mutator Chooser** was *SPM4i*, the **Sampler** used *GROMACS* topologies, the **Scorers** were *EvoEF2*, *BLUUES*, *PIEPISA* and *gmxMMPBSA* and the *Consensus Threshold* **Pruner** was used to select the best candidates. After the protocol run, all average scores had at least doubled and many promising candidates were recovered from the last epoch. These can then be further validated by longer MD simulations, rescored if necessary, and, finally, undergo experimental validation.

This example, as well as further details on the specific features of each **Block**, can be found at https://lo cuaz.readthedocs.io/en/latest/tutorialsimple.html

Further tutorials are available in the documentation, each of them providing optimal parameters for different use cases: using topologies from GROMACS, building the topology from Tleap and optimizing a protein against a small ligand.

Acknowledgements

PB acknowledges the TRIL fellowship for having provided support under the ICTP TRIL programme, Trieste, Italy. We acknowledge PRACE for awarding us access to the supercomputing resources hosted by CINECA and the VEGA resources hosted at the Institute of Information Science, IZUM.

Conflict of interest

No competing interest is declared.

Funding

The research leading to these results has received funding from AIRC under IG 2020 - project ID. 24589.

References

Abraham MJ, Murtola T, Schulz R *et al.* Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 2015;**1-2**:19–25.

Adolf-Bryfogle J, Kalyuzhniy O, Kubitz M *et al.* RosettaAntibodyDesign (RAbD): A general framework for computational antibody design. *PLoS Comput Biol* 2018;**14**:E1006112. https://doi.org/10.1371/journal.pcbi.1006112

Barnes JE, Lund-Andersen PK, Patel JS *et al.* The effect of mutations on binding interactions between the SARS-CoV-2 receptor binding domain and neutralizing antibodies B38 and CB6. *Sci Rep* 2022;**12**: 18819–2045. https://doi.org/10.1038/s41598-022-23482-5

Buratto D, Wan Y, Shi X *et al.* In silico maturation of a nanomolar antibody against the human cxcr2. *Biomolecules* 2022;**12**:1285.

D'Agostino S, Mazzega E, Praček K *et al.* Interference of p53: twist1 interaction through competing nanobodies. *Int J Biol Macromol* 2022;**194**:24–31. https://doi.org/10.1016/j.ijbiomac.2021.11.160

Hallen MA, Martin JW, Ojewole A *et al.* Osprey 3.0: open-source protein redesign for you, with powerful new features. *J Comput Chem* 2018;**39**:2494–507. https://doi.org/10.1002/jcc.25522.

Hie BLL, Shanker VRR, Xu D *et al.* Efficient evolution of human antibodies from general protein language models. *Nat Biotechnol* 2023; **42**:275–83. https://doi.org/10.1038/s41587-023-01763-2

Holt GT, Gorman J, Wang S *et al.* Improved hiv-1 neutralization breadth and potency of v2-apex antibodies by in silico design. *Cell Rep* 2023; **42**:112711. https://doi.org/10.1016/j.celrep.2023.112711

Kennedy PJ, Oliveira C, Granja PL *et al.* Monoclonal antibodies: technologies for early discovery and engineering. *Crit Rev Biotechnol* 2018;**38**:394–408.

Mason DM, Friedensohn S, Weber CR *et al.* Optimization of therapeutic antibodies by predicting antigen specificity from antibody sequence via deep learning. *Nat Biomed Eng* 2021;**5**:600–12.

Moriyama S, Anraku Y, Taminishi S *et al.* Structural delineation and computational design of sars-cov-2-neutralizing antibodies against omicron subvariants. *Nat Commun* 2023;**14**:4198. https://doi.org/10.1038/s41467-023-39890-8

Ochoa R, Soler MA, Laio A *et al.* PARCE: Protocol for amino acid refinement through computational evolution. *Computer Physics Communications* 2021;**260**:107716–00104655. https://doi.org/10.1016/j.cpc.2020.107716

Salomon-Ferrer R, Case DA, Walker RC. An overview of the amber biomolecular simulation package. *WIREs Comput Mol Sci* 2013;**3**:198–210. https://doi.org/10.1002/wcms.1121

Shan S, Luo S, Yang Z *et al.* Deep learning guided optimization of human antibody against sars-cov-2 variants with broad neutralization. *Proc Natl Acad Sci U S A* 2022;**119**:E 2122954119. https://doi.org/10.1073/pnas.2122954119

Shanehsazzadeh A, Bachas S, McPartlon M *et al.* Unlocking *de novo* antibody design with generative artificial intelligence. *Synth Biol* 2023, Preprint: not peer reviewed.

Sivasubramanian A, Sircar A, Chaudhury S *et al.* Toward high-resolution homology modeling of antibody F_v regions and application to antibody-antigen docking. *Prot Struct Funct Bioinform* 2009;**74**:497–514. https://doi.org/10.1002/prot.22309

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**:2740–8. https://doi.org/10.1039/C6CP07807A

Soler MA, Fortuna S, de Marco A *et al.* Binding affinity prediction of nanobody–protein complexes by scoring of molecular dynamics trajectories. *Phys Chem Chem Phys* 2018;**20**:3438–44. https://doi.org/10.1039/C7CP08116B

Soler MA, Medagli B, Semrau MS *et al.* A consensus protocol for the in silico optimisation of antibody fragments. *Chem Commun (Camb)* 2019;**55**:14043–6. https://doi.org/10.1039/C9CC06182G

Soler MA, Minovski N, Rocchia W *et al.* Replica-exchange optimization of antibody fragments. *Comput Biol Chem* 2023;**103**:107819.

Sormanni P, Aprile FA, Vendruscolo M. Third generation antibody discovery methods: In silico rational design. *Chem Soc Rev* 2018;**47**:9137–57.

Tandiana R, Barletta GP, Soler MA *et al.* Computational mutagenesis of antibody fragments: disentangling side chains from $\Delta\Delta G$ Predictions. *JCTC* 2024;**20**:2630–42. https://pubs.acs.org/doi/10.1021/acs.jctc.3c01225

Valdés-Tresanco MS, Valdés-Tresanco ME, Valiente PA *et al.* Gmx_MMPBSA: a new tool to perform End-State free energy calculations with GROMACS. *J Chem Theory Comput* 2021;**17**:6281–91. https://doi.org/10.1021/acs.jctc.1c00645

Bioinformatics, 2024, 40, 1–4
https://doi.org/10.1093/bioinformatics/btae492
Applications Note