

## RESEARCH ARTICLE

## RosettaSurf—A surface-centric computational design approach

Andreas Scheck<sup>1,2</sup>, Stéphane Rosset<sup>1,2</sup>, Michaël Defferrard<sup>3</sup>, Andreas Loukas<sup>3</sup>,  
Jaume Bonet<sup>1,2</sup>, Pierre Vandergheynst<sup>3</sup>, Bruno E. Correia<sup>1,2\*</sup>

**1** Institute of Bioengineering, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, **2** Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland, **3** Signal Processing Laboratory (LTS2), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

\* [bruno.correia@epfl.ch](mailto:bruno.correia@epfl.ch)



## OPEN ACCESS

**Citation:** Scheck A, Rosset S, Defferrard M, Loukas A, Bonet J, Vandergheynst P, et al. (2022) RosettaSurf—A surface-centric computational design approach. *PLoS Comput Biol* 18(3): e1009178. <https://doi.org/10.1371/journal.pcbi.1009178>

**Editor:** Joanna Slusky, University of Kansas, UNITED STATES

**Received:** June 14, 2021

**Accepted:** February 21, 2022

**Published:** March 16, 2022

**Copyright:** © 2022 Scheck et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All data and scripts necessary to recreate the analysis and design simulations described in this work are available at <https://github.com/LPDI-EPFL/RosettaSurf>.

**Funding:** B. E. C. was supported by the European Research Council (starting grant no. 716058), the Swiss National Science Foundation (grant no. 310030\_188744), the NCCR Molecular Systems Engineering ([www.nccr-mse.ch](http://www.nccr-mse.ch)) and the NCCR Chemical Biology ([www.nccr-chembio.ch](http://www.nccr-chembio.ch)). A. L. was supported by the Swiss National Science

## Abstract

Proteins are typically represented by discrete atomic coordinates providing an accessible framework to describe different conformations. However, in some fields proteins are more accurately represented as near-continuous surfaces, as these are imprinted with geometric (shape) and chemical (electrostatics) features of the underlying protein structure. Protein surfaces are dependent on their chemical composition and, ultimately determine protein function, acting as the interface that engages in interactions with other molecules. In the past, such representations were utilized to compare protein structures on global and local scales and have shed light on functional properties of proteins. Here we describe RosettaSurf, a surface-centric computational design protocol, that focuses on the molecular surface shape and electrostatic properties as means for protein engineering, offering a unique approach for the design of proteins and their functions. The RosettaSurf protocol combines the explicit optimization of molecular surface features with a global scoring function during the sequence design process, diverging from the typical design approaches that rely solely on an energy scoring function. With this computational approach, we attempt to address a fundamental problem in protein design related to the design of functional sites in proteins, even when structurally similar templates are absent in the characterized structural repertoire. Surface-centric design exploits the premise that molecular surfaces are, to a certain extent, independent of the underlying sequence and backbone configuration, meaning that different sequences in different proteins may present similar surfaces. We benchmarked RosettaSurf on various sequence recovery datasets and showcased its design capabilities by generating epitope mimics that were biochemically validated. Overall, our results indicate that the explicit optimization of surface features may lead to new routes for the design of functional proteins.

## Author summary

Finely orchestrated protein-protein interactions are at the heart of virtually all fundamental cellular processes. Altering these processes or encoding new functions in proteins has

Foundation project “Deep Learning for Graph-Structured Data” (grant number PZ00P2 179981). J. B. was funded by the EPFL Fellows postdoctoral fellowship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

been a long-standing goal in computational protein design. Protein design methods commonly rely on scoring functions that seek to identify amino acid sequences that optimize structural configurations of atoms while minimizing a variety of physics-based and statistical terms. The objectives of the large majority of computational design protocols have been focused on obtaining a predefined structural conformation. However, routinely introducing a functional aspect on designer proteins has been more challenging. Our results suggest that the molecular surface features can be a useful optimization parameter to guide the design process towards functional surfaces that mimic known protein binding sites and interact with their intended targets. Specifically, we demonstrate that our design method can optimize experimental libraries through computational screening, creating a basis for highly specific protein binders, as well as design a potent immunogen that engages with site-specific antibodies. The ability to create proteins with novel functions will be transformative for biomedical applications, providing many opportunities for the design of novel immunogens, protein components for synthetic biology, and other protein-based biotechnologies.

## Introduction

Proteins are key components in living cells, performing many functions that commonly rely on physical interactions between molecules. The molecular surface arising from the three-dimensional arrangement of the many atoms that compose a protein is determinant for protein function and is therefore crucial for biological processes [1]. While discrete atomic-level protein representations have been invaluable for our understanding of protein function, near-continuous surface-based representations offer the opportunity to study protein structures using a different subset of features (e.g. electrostatic potentials, geometry).

In 1971, Lee and Richards introduced the concept of solvent-accessible surfaces, which in practice are generated by rolling a probe approximating a solvent molecule over the protein atoms [2]. The molecular surface, often denoted as solvent-excluded surface or Connolly surface [3–5], consists of the surface that can be directly contacted by the probe and the reentrant surface which smooths over gaps between the atoms that were not accessible to the probe sphere [4,6,7]. Numerical representations of surfaces have also been developed, ranging from dot surfaces, to voxel representations and graphs [1,8–12]. These representations allow the mapping of molecular and geometric properties onto the generated surface, including physico-chemical properties (such as electrostatics and hydrophobicity), and geometrical features (e.g. protrusions or cavities) [6,8,13–15].

Intuitively, the molecular surface forms the boundary of the protein and its surroundings, thus acting as the interface that engages in interactions with other molecules. The study of protein structures as near-continuous molecular surfaces is therefore important to understand structural and functional aspects of proteins, which may not be fully captured by a discrete atomic representation [13,16]. A widely studied category of features of molecular surfaces is their chemistry, in particular electrostatic potentials, and their implications for protein function. Most notably, such representations have been used to study various types of protein-substrate interactions [17,18]. Other important metrics used to study molecular surfaces are their shape-derived properties, commonly focusing on shape complementarity ( $S_C$ ) or shape similarity ( $S_S$ ). Shape complementarity is frequent in the study of molecular recognition, e.g. protein-protein or protein-ligand interactions. In particular, protein-protein interactions (PPIs) have been extensively studied in terms of complementarity, showing that protein-protein

interfaces are often highly complementary, both in shape and in charge [19,20]. On the other hand, shape similarity has been used to globally and locally compare biomolecules, aiding the functional annotation of proteins that show structural similarities but lack detectable sequence homology [11,21].

An important extension of such successful applications in analysis and prediction tasks using surface-based representations is that of protein design [6], where the objective is to guide the sequence search process to optimize specified surface features.

Here we present a new, surface-centric computational design strategy, termed RosettaSurf, that uses a description of the molecular surface shape and electrostatic properties as objective function for scoring optimization in protein design. Working at the surface-level of protein structures offers a unique approach for the design of proteins and their functions, as molecular surfaces are, to some extent, independent of the underlying sequence [11,14]. This means that different sequences in different proteins may present similar surfaces [11,14]—a premise that drives our proposed methodological approach. In contrast, the large majority of computational design workflows entail a discrete atomistic description of the proteins and sequence design is typically performed on the atom arrangements and interactions; i.e. by sampling different amino acid side chains and adjusting atoms in the protein structure to optimize a given energy function.

While these strategies have been successfully applied to design novel protein topologies [22–24], the design of functional proteins based solely on computational calculations remains a challenge [25–32].

RosettaSurf is a protocol implemented in the Rosetta software package [33] and we assessed its performance with several benchmark tests, demonstrating the protocol's capabilities through the recovery of protein interfaces and its application to functional protein design.

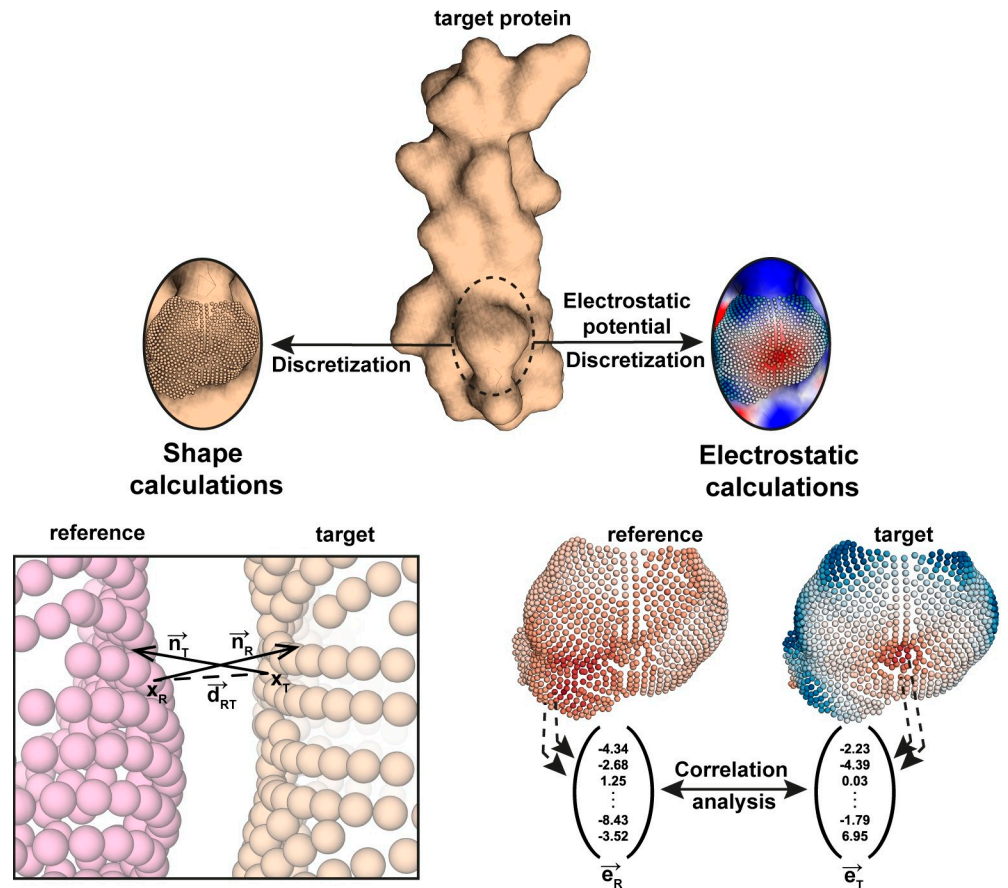
## Results

### RosettaSurf framework

The RosettaSurf protocol operates at the solvent-excluded surface level of a protein structure, and its core operation is the comparison between surfaces. To compare molecular surfaces, we defined a score that quantifies the similarity between two surfaces considering both shape and electrostatic features, and incorporated it within the Rosetta software package. The molecular surface is generated from the three-dimensional atomic coordinates of a protein [3,4] and stored as a discrete point cloud (Fig 1). Representation of the surface as a point cloud allows the featurization of the points with geometrical and chemical descriptors (Fig 1) and enables rapid calculations of the surface similarity score. We refer to the mutable surface as target while the reference surface used for comparisons is denoted as reference.

To describe the surface geometry, we developed a descriptor based on concepts introduced by Lawrence and Colman [19] that quantifies shape relationships of two surfaces relying on normal vector comparisons that are derived at each point of the surface (Fig 1). For all pairs of closest points of the target and reference surface the geometric similarity is evaluated. Averaging these individual similarity values yields an overall shape similarity ( $S_S$ ) score that quantifies the similarity of the geometry of two given molecular surfaces. The  $S_S$  score ranges from 0 (no similarity) to 1 (identical shape).

Electrostatic similarity ( $E_S$ ) of two surfaces is derived from comparing their electrostatic potentials, i.e.  $E_S$  is assessed by computing the correlations of the two electrostatic fields originating from the target and reference surface, respectively (Fig 1). Correlation coefficients are derived from this comparison that describe the similarity of the electrostatic fields. Values of  $E_S$  range from -1 (highly similar) to +1 (highly complementarity).



**Fig 1. Computation of the surface similarity score ( $\text{Surf}_s$ ).** Protein surface is converted into a point cloud where each point is used to compute shape and electrostatic features. To compute the  $\text{Surf}_s$  score, all individual points of the point clouds are compared and the shape similarity value is derived from closest points of the two surfaces in space while electrostatic similarity is evaluated by correlation analysis of the electrostatic potentials of both surfaces.

<https://doi.org/10.1371/journal.pcbi.1009178.g001>

To accurately capture the relative contributions of both shape and electrostatics on the similarity of molecular surfaces, we combined both scores into a single surface similarity score ( $\text{Surf}_s$ ). Since we were interested in assessing similarity of surface sites that are well maintained and relevant to functional protein design, we focused on interfaces of PPIs and performed logistic regression on a dataset of 2,660 protein complexes to optimize the individual weights of the  $\text{Surf}_s$  score for each component, shape and electrostatics.

The  $\text{Surf}_s$  score is defined as follows:

$$\text{Surf}_s = \frac{1}{1 + \exp(13.79986756 - 14.64347448 \cdot S_s + 13.78594078 \cdot E_s)}$$

The final  $\text{Surf}_s$  score combines both properties into a single score scaled from 0 to 1, where 1 represents highly similar surfaces. A detailed explanation of the computation of the individual scores and the resulting  $\text{Surf}_s$  score can be found in Materials and Methods. We note that the developed surface similarity score can be straightforwardly converted to surface complementarity ( $S_c$ ) by inverting the normal vectors of the reference surface. Complementarity has been typically used to evaluate PPIs and was first demonstrated by Lawrence and Coleman for molecular shapes [19].

## Surface-centric protein design protocol

We developed the RosettaSurf design protocol, as part of the RosettaScript software environment [34], which utilizes the described surface scoring function during the sequence sampling stage of the design process to bias the selection of amino acids and rotamers towards a desired surface configuration in terms of geometric and electrostatic properties (S1 Fig). Thereby, it is possible to design proteins for function without relying on protein grafting but by optimizing the molecular surface, for instance, to mimic a given active site.

With the RosettaSurf design protocol we explicitly optimize molecular surface features during the protein design process. In practice, RosettaSurf performs sequence design in a way where the mutable surface of the design scaffold (target) is optimized to closely match the surface features of a reference protein (reference) (S1 Fig). To efficiently explore the sequence space during the design process, Monte Carlo simulated annealing guides the optimization of rotamers, where substitutions of residues are scored based on the resulting surface and accepted if they pass the Monte Carlo criterion that is implemented as the Surf<sub>s</sub> score.

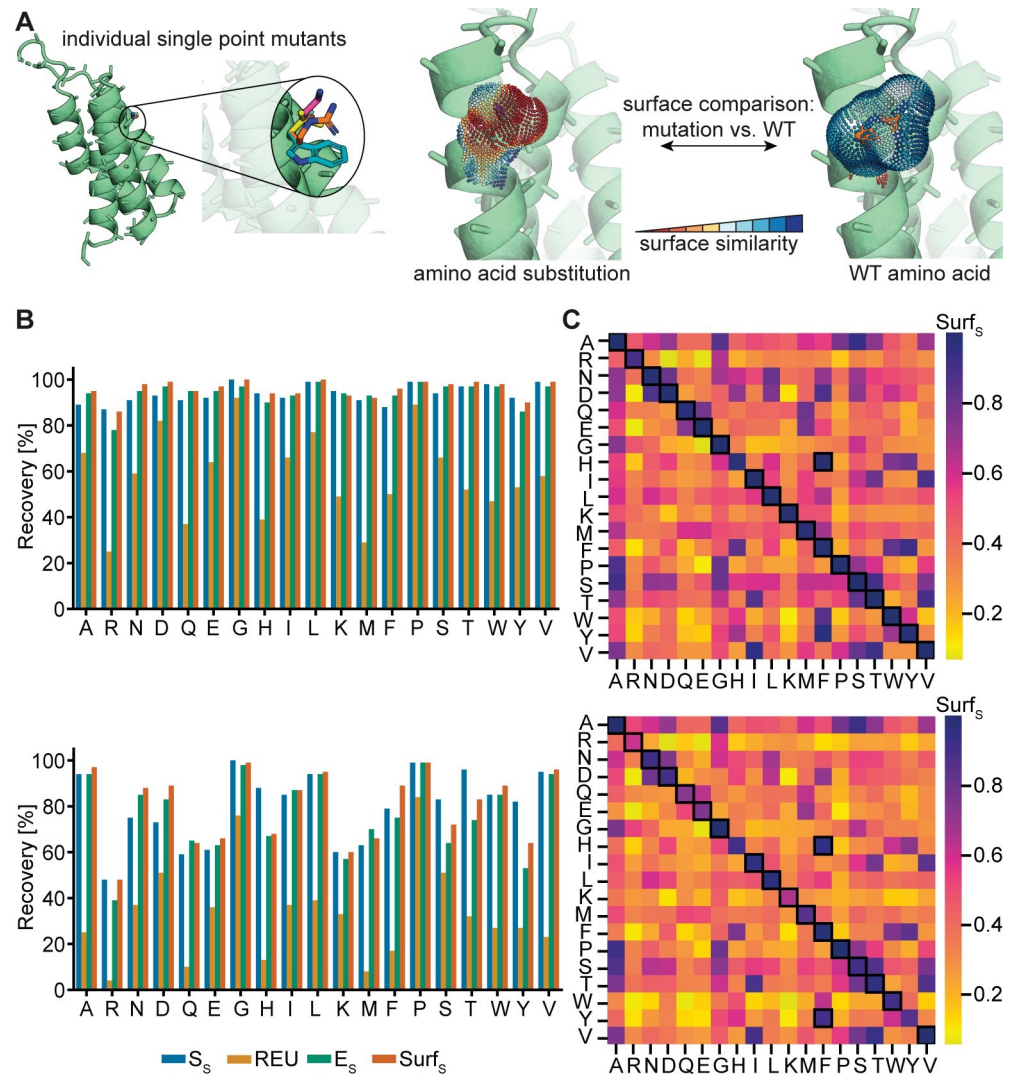
To reduce computational time spent on rotamer sampling in a combinatorial fashion on the overall surface, we implemented a single amino acid scanning surface-centric protein design approach (RosettaSurf-site). This protocol samples amino acids individually at each position of the design surface, selects the top three rotamers according to the surface and samples those combinatorially with other designable positions, reducing the combinatorial possibilities.

In our benchmark studies we sought to compare the developed surface-centric design protocol, RosettaSurf, to state-of-the-art macromolecular design approaches implemented in Rosetta. The Rosetta energy function has been parametrized to evaluate and optimize the energy of many different aspects of molecular interactions (e.g. protein stability, protein-ligand, protein-protein and protein-nucleic acids, etc.); it contains both statistical and physics-based terms, being calibrated using a discrete atomic representation of the molecules. This type of optimization has been particularly successful for the design of novel sequences that fold into defined protein structures. However, to design proteins that display defined motifs which can perform biological functions has proven to be difficult. By focusing the design process on the areas where molecular interactions occur—the protein surfaces—novel approaches attempting to design function into proteins may represent new routes to address this problem.

Two types of metrics in surface comparison are considered: surface complementarity and surface similarity. We show how the surface similarity score captures features of individual residues and demonstrate the ability of the RosettaSurf protocol to recover amino acid sequences of native protein interfaces. Furthermore, we highlight the performance of RosettaSurf for the design of surface patches at protein interfaces.

## Single amino acid recovery

To evaluate the accuracy of our design protocol, we performed a benchmark on the recovery of single amino acids in native protein interfaces. The amino acid of interest is substituted by each of the 19 amino acid identities, excluding cysteine as it can form covalent bonds in the form of disulfide bridges, thus rendering the resulting surface dependent on the interplay of two amino acids, and the surfaces of the substituted amino acids are compared to the surface of the native rotamer (Fig 2A). These calculations are performed without the knowledge of the native amino acid surface and this information is only used for the assessment of similarity upon each substitution. For each amino acid type a dataset of 100 crystal structures of protein complexes forming transient interactions was compiled, resulting in a total of 1,900 complexes (for details see [Materials and Methods](#)). The energy computed by Rosetta, Rosetta Energy Unit



**Fig 2. Single mutant discrimination using surface similarity score in protein-protein complexes.** A) Surface similarity evaluation protocol for single amino acids. B) Recovery for all 19 considered amino acid types in bound (top) and unbound (bottom) complex states, evaluated with four different metrics:  $S_s$  (shape similarity), REU (Rosetta energy unit),  $E_s$  (electrostatic similarity), and  $Surf_s$  (surface similarity). C) Average surface similarity score when performing all-against-all amino acid comparison for bound (top) and unbound (bottom) complex states. The highest mean  $Surf_s$  score for every amino acid is highlighted.

<https://doi.org/10.1371/journal.pcbi.1009178.g002>

(REU), serves as a baseline for comparison to the surface-centric design strategy. We evaluated the recovery of the different amino acids for mutations made in the bound and unbound state of the protein complex, respectively (Fig 2B). The recovery of an amino acid is deemed successful when the native amino acid has the best score among all 19 amino acids and is uniquely identifiable. In this benchmark we evaluated the performance of shape similarity, electrostatic similarity and both combined in a surface similarity score. The surface similarity measurement is highly accurate in identifying native amino acids in the bound state of the protein complexes, showing consistently higher sequence recovery rates than Rosetta energy (Fig 2B).

Incorporating the electrostatic similarity term into the surface score generally results in a boost in recovery rates over shape similarity alone, in particular for amino acids that have

close structural doubles but differ in chemical properties, e.g. glutamine and glutamate, as well as, asparagine and aspartate (Fig 2B).

The Rosetta energy function shows the best recovery rates for amino acids with unique features on their side-chains, i.e. glycine and proline. However, even for these cases, surface similarity outperforms the recovery by REU. A similar trend can be observed for the unbound complexes, although the general success of retrieving the native amino acid decreases for certain amino acids. This trend is most obvious for polar residues and large side chains (e.g. arginine, lysine, or methionine) that have access to a variety of different rotamer conformations as the lack of the binding partner allows a larger conformational space for different rotamers to be explored. For these types of amino acids, the sequence recovery rate is higher in the bound conformation as the presence of the binder restricts the possible rotamer space that can be explored, especially for highly exposed residues (S2A and S2B Fig). A similar degradation in performance can be observed for tyrosine, phenylalanine, and histidine in the unbound benchmark case. These amino acids share common surface signatures according to our score (Fig 2C), complicating their recovery which is amplified in the unbound benchmark cases.

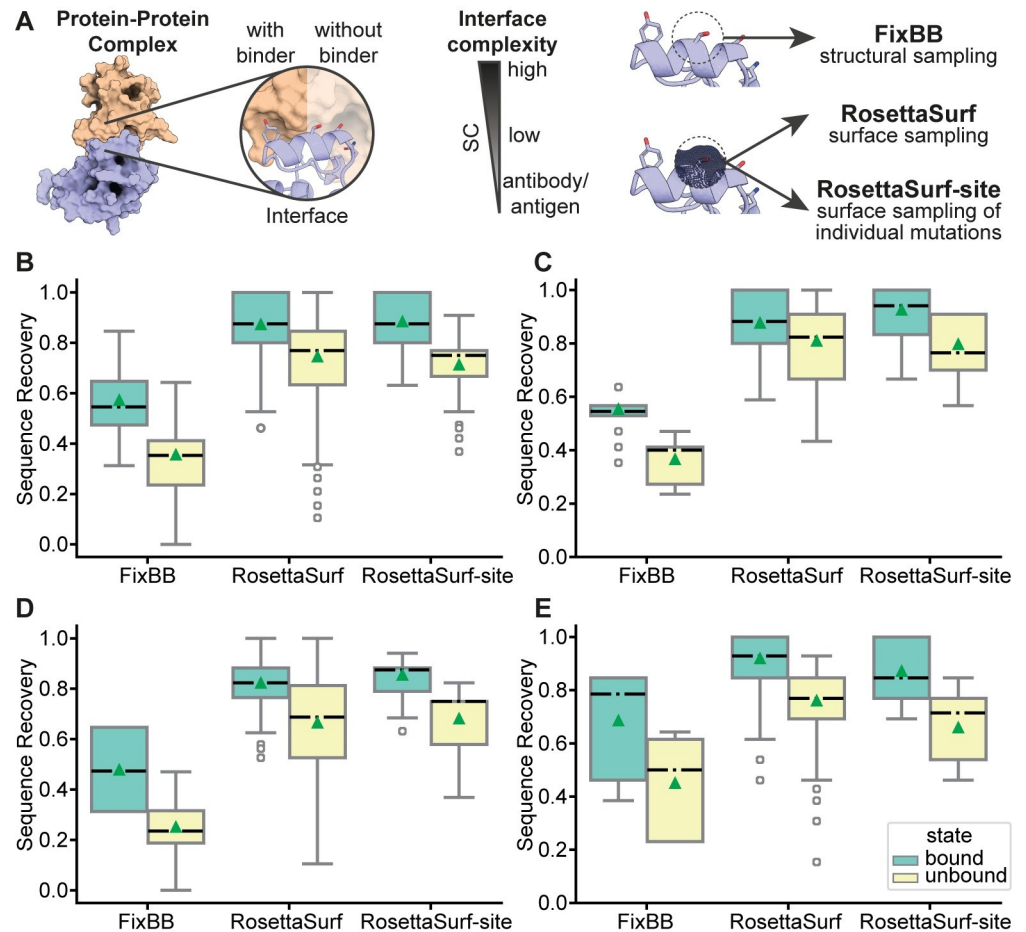
Moreover, the surface similarity scores retrieved from an all-against-all amino acid comparison demonstrate the high accuracy by which most amino acids can be identified (Fig 2C). Here, we computed mean surface similarity values for all substitutions tested, i.e. for each native amino acid across the 100 protein complexes we computed the similarity of all other 19 amino acids to the native one on average. The results show that each of the 19 amino acids is generally most similar to itself, demonstrating that the method can accurately distinguish between the different amino acid types. Notable exceptions are residues that share similar geometrical features, e.g. phenylalanine and histidine or tyrosine.

Close inspection of the considered surface of a single amino acid demonstrates the local precision of the surface similarity score (S3 Fig). Small off-rotamer deviations are captured on the point cloud and are specific for the mismatching surface area. These results indicate that the surface similarity score is sensitive to small differences in the comparison of two surface point clouds. Overall, these results clearly show that the developed surface similarity score can capture local differences in surfaces providing the basis for the evaluation of differences between full surface patches.

## Protein interface sequence recovery

Having shown that the implemented surface similarity score is sufficiently accurate to recover individual amino acids, the following benchmark expands on recovering entire surface patches in protein interfaces. We evaluated the ability of our surface-based sequence design protocols (RosettaSurf and RosettaSurf-site) to recover natural protein interfaces as a gateway for the design of proteins endowed with biochemical function. By focusing on protein interfaces, the correct sequence does not solely depend on minimizing the Rosetta scoring function, but rather needs to represent the surface properties of the interface site. While such design scenario has limited applicability for the *de novo* design of protein-protein interactions, it is important for applications in the domain of immunogen design for vaccine development where the surface mimicry of known surfaces (neutralizing epitopes) is critical for the biological activity of this type of design [28,30–32].

In this benchmark we considered nine protein-protein complexes forming transient interactions (for details see [Materials and Methods](#)), grouped into three categories (low complementarity, high complementarity, antibody/antigen), and evaluated the performance by assessing sequence recovery (Fig 3A). We compared three different design approaches:



**Fig 3. Sequence recovery of protein interfaces.** A) Sequence recovery benchmark pipeline. Sequences in the interfaces of protein-protein complexes are evaluated in the presence and absence of the binding partner. The tested complexes were grouped into interfaces with low and high shape complementarity, and antibody-antigen complexes. Surface-centric design (RosettaSurf and RosettaSurf-site) is compared to a standard structural protein design protocol (FixBB). B) Interface sequence recovery of the complete dataset. C) Sequence recovery of low shape complementarity interfaces. D) Sequence recovery of high shape complementarity interfaces. E) Sequence recovery of antigen-antibody complexes. Dashed lines represent median and triangles represent mean recovery values.

<https://doi.org/10.1371/journal.pcbi.1009178.g003>

1. Standard Rosetta design (FixBB)
2. Surface-centric design (RosettaSurf)
3. Single-site-scanning surface-centric design (RosettaSurf-site)

All protocols operate on fixed backbones of the protein complexes that are presented in the native conformation that mediates the interaction. Consequently, the sequence recovery success will largely depend on side-chain placement in a given backbone. The surface-centric design approaches are compared to the standard FixBB Rosetta design protocol, using the ref2015 scoring function [35]. In the presented benchmarks, the Rosetta FixBB design protocol serves as baseline to assess the impact of surface-centric design on sequence recovery.

In a first step, the target protein's interface is stripped off its native sequence by mutating all interface residues to alanine. Second, the different design protocols were employed on the interface positions with FixBB using the Rosetta energy function to select mutations while



RosettaSurf was guided by our surface similarity scoring term (Fig 3A). RosettaSurf was therefore provided with the native protein surface that we aimed to mimic as reference surface, thus actively optimizing towards that given native surface during the design process. With this setup it is possible to evaluate the sensitivity of the Surf<sub>s</sub> score to recover native amino acids compared to a known ground truth. This is in contrast to the single amino acid recovery benchmark where the reference surface solely served as comparison to evaluate the similarity of the native surface and the surface resulting from the mutation without affecting the amino acid change itself.

All different design protocols were employed for the bound and unbound states of the protein complexes and results are reported by their category as mean sequence recovery rate (Fig 3A). Overall, sequence recovery rates were higher for surface-centric design pipelines while standard Rosetta sequence design showed lower recovery rates regardless of the presence of the binder (Fig 3B). For all studied complexes, RosettaSurf and RosettaSurf-site outperformed FixBB by 36–39 percentage points in the unbound and 30–31 percentage points in the bound state. As expected, sequence recovery was in general more successful in the presence of the binding partner due to the reduced number of possible side-chains and rotameric conformations. However, even in the presence of the binder, surface-centric design can be applied to improve results. FixBB demonstrated improved performance for the antibody/antigen test set, however, the surface-centric protocols still reached better results that were comparable to the other categories. Worth noting is the 100% sequence recovery success of RosettaSurf-site for the D8 protein-vv138 antibody complex. The FixBB protocol performed slightly better for low-complementary (55%) than for high-complementarity complexes (48%), while RosettaSurf and RosettaSurf-site resulted in similar recovery rates of ~85% in both cases with the notable exception of RosettaSurf-site for low-complementarity complexes. Here, RosettaSurf-site was able to recover 92% of the sequence and for one complex of that category, the Enterotoxin G–T-cell receptor complex, even 100%.

Furthermore, we analyzed more closely outlier decoys, i.e. structures that scored high in surface similarity but showed only little sequence recovery when designing with RosettaSurf. We set a threshold of up to 50% as low recovery and a Surf<sub>s</sub> score of greater or equal to 0.9992 as this value marks the lower end of structures with high sequence recovery ( $> = 70\%$ ), leading to one candidate from the RSV epitope-scaffold-Motavizumab complex. When investigating which types of amino acids were common failures, we identified general geometric resemblance as likely reason. Several amino acids are frequently replaced by structurally similar residues, e.g. isoleucine by valine and lysine by arginine.

As for sequence recovery without the binder, RosettaSurf and RosettaSurf-site performed better for complexes of low shape complementarity with 81% and 80% recovery, respectively. However, sequence recovery obtained for high-complementary interfaces reached 66% for RosettaSurf and 68% for RosettaSurf-site. For antibody/antigen complexes we observed 76% for RosettaSurf and 66% for RosettaSurf-site. All these performances were substantially higher than those of FixBB, that reached 25% and 45% for high-complementary and antibody/antigen complexes, respectively.

In addition, we investigated possible reasons for structures with low sequence recovery but high surface similarity scores. We adjusted the selecting criteria to structures with  $< = 30\%$  sequence recovery and  $> = 0.9996$  Surf<sub>s</sub> score, observing similar results as for the bound benchmark. Four structures, all from the Colicin E9-IM9 complex fulfill these criteria. Again, mainly amino acids with similar geometrical features were common mismatches between recovered and native amino acids, e.g. phenylalanine, tyrosine, and histidine, aspartic acid and asparagine, as well as valine and threonine.

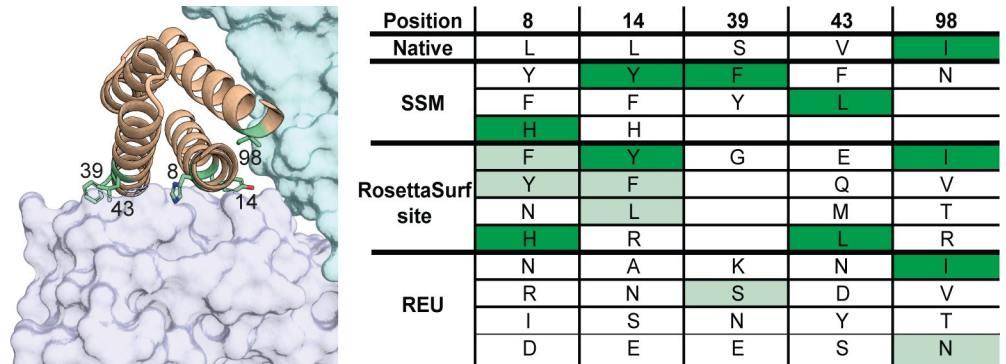
Overall, similar performance of RosettaSurf and RosettaSurf-site showed that performing the combinatorial sampling of all possible rotamers simultaneously was not necessary. During the RosettaSurf-site pipeline amino acids were sampled individually at each design position and only a subset of the amino acids was used for combinatorial sampling based on the surface similarity score, resulting in similar recovery outcomes. Furthermore, despite small differences in the sequence recovery rate between the three selected categories of protein complexes, overall no strong correlation between the shape complementarity of the considered complexes and the success in sequence recovery could be observed. We considered groupings based on different metrics, namely total interface area, hydrophobic interface area, and number of recoverable positions. No correlation with sequence recovery was observed for these different properties (S4 Fig). Additionally, we grouped the complexes based on their proportion of amino acids with access to diverse rotamer conformations to the total number of residue positions considered during the recovery. A weak correlation could be observed between reduced recovery rate and higher numbers of flexible amino acids (S4 Fig), in line with the results from the single amino acid recovery benchmark (Fig 2B). Together, these results suggest that the proposed method is largely independent of the provided protein structure.

Our results have thus implications for the design of functional proteins as the high success of recovering natural protein interfaces may be promising for the design of proteins displaying defined surface properties, as is the case for immunogen design.

### Computational SSM screening to improve protein binding

A common problem in protein design is the optimization of PPIs to generate high affinity protein binders. Experimental screening methods, e.g. site-saturation mutagenesis (SSM) and combinatorial libraries, provide insights into mutations that improve binding interactions but are time and resource-consuming experiments. While *in silico* saturation mutagenesis has been used previously to identify stabilizing mutations or improve binding affinity, these approaches were based on energy functions to identify beneficial mutations [36–40]. However, to our knowledge, surface metrics have not been considered as sole selection criterion and compared to experimentally determined mutations. Based on our observations in the sequence recovery benchmarks, we tested if RosettaSurf-site could be a fast and efficient computational screening alternative to experimental based approaches. We studied the optimization of *de novo* designed interleukin-2/15 antagonists that bind specifically to the IL-2R $\beta\gamma_c$  receptor [41]. The computationally designed interleukins were optimized for binding to the interleukin receptor by performing SSM followed by combinatorial libraries based on the identified mutations. The described study did not only report sequence information but included additional structural characterization of the resulting design, thus allowing a fair comparison to our computational method. This level of data completeness is rare in many other design endeavors that have used this type of optimization strategy [42,43].

We performed RosettaSurf-site similar to the approach described above (see “Protein Interface Sequence Recovery”), however, with the selection criteria being shape complementarity of the design and IL-2R $\beta\gamma_c$  interface rather than similarity as shape complementarity has been shown to be a key feature of high affinity binding interactions [44]. RosettaSurf-site allows exhaustive computational sampling of amino acids similar to SSM experiments, screening all possible amino acid substitutions at the interface of the interleukin design. Our study is based on the reported crystal structure of the interleukin-2/15 design in complex with its cognate receptor (PDB: 6DG5). We selected interface positions that were tested during SSM and later included in the combinatorial library for our benchmark screening. Positions not interacting with the receptor, i.e. not within a  $C_\beta$ -distance of 7 Å, were excluded from the selection,



**Fig 4. Comparison of SSM data obtained for the designed interleukin-2/15 antagonists in comparison to RosettaSurf-site predictions.** The structure highlights the five selected positions of the interleukin design that were computationally and experimentally sampled. Different sampling results of the experimental SSM, RosettaSurf-site with SC, and Rosetta's energy function are reported in the table. Mutations resulting in the experimentally reported best binding design are highlighted in dark green. RosettaSurf-site was able to recover four out of five key binding mutations (dark green) while evaluating mutations with Rosetta's energy function could only retrieve one binding mutation at position 98. Additionally, RosettaSurf-site was able to recover four affinity improving mutations not present in the best binding design, whereas Rosetta's energy function could identify only two of these mutations (light green).

<https://doi.org/10.1371/journal.pcbi.1009178.g004>

resulting in a total of 18 positions that can be regarded as interface. From these 18 possible interface residues, four positions were also present in the experimental library. One additional position (residue 39) was included that lies at the boundary of the interface, within 9 Å of the receptor, and was tested in the experimental library, resulting in a total of five positions as design input for our benchmark. In our design protocol, the five selected residues were sampled with RosettaSurf-site and mutations were evaluated based on the shape complementarity to the receptor surface. The computational results were compared to the composition of the combinatorial library that was constructed based on the preceding SSM screening. We analyzed whether RosettaSurf-site was able to identify mutations that are present in the highest affinity binder as well as the performance of the Rosetta energy function to retrieve similar mutations. After performing all possible mutations with RosettaSurf-site, the four top-ranking amino acids for each position were selected (Fig 4), in line with the maximum number of amino acids included in the combinatorial library.

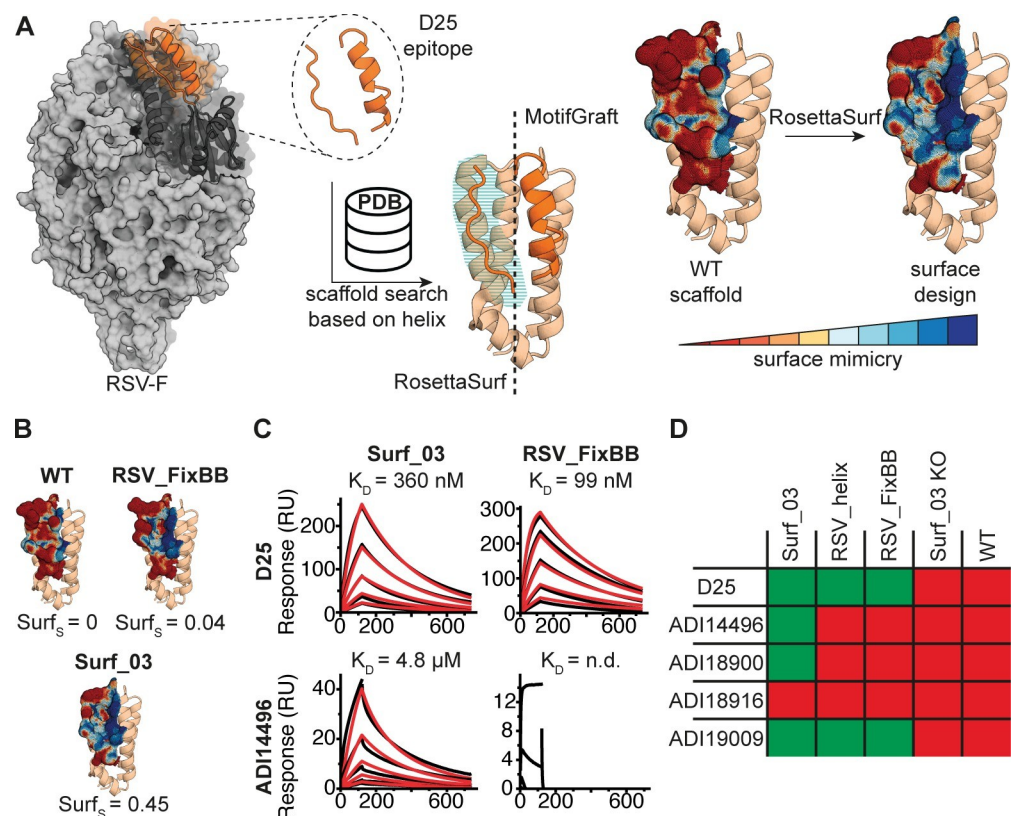
RosettaSurf-site recovered four out of the five mutations present in the best design, with residue 39 being at the edge of the interface and contributing minimally to the protein-protein interaction, thus making it more challenging for this approach. Four additional residues included in the combinatorial library were recovered, that were observed to improve binding in the SSM but were not present in the highest affinity binder. In contrast, selecting mutations based on Rosetta energy alone recovers only a single mutation observed in the best binder and two residues present in the combinatorial library. These results show the potential of surface-based design of single point mutants as fast and promising approach to select candidates for combinatorial libraries to improve binding interactions without requiring preceding screening experiments like SSM.

### Surface-centric design of a novel RSV site 0 immunogen

In recent years, protein design has shown promising results in the field of immunoengineering, allowing the computational design of epitope-focused immunogens that were shown to elicit functional antibody responses in mice and non-human primates [28,30,31,45]. To achieve an epitope-specific immune response, the epitope is transplanted from the viral

antigen to an unrelated small protein scaffold, presenting the antigenic site in isolation. We applied the RosettaSurf protocol to the design of immunogens mimicking the antigenic site 0, an epitope on the F-glycoprotein of Respiratory Syncytial Virus (RSV) which consists of an irregular  $\alpha$ -helix and a 10-residue long loop. Previous studies have shown that identifying scaffolds that can present such complex structural motifs is challenging, and specifically for site 0 they are currently not available in the known structure space. *De novo* design methods have been successfully used to build scaffolds from scratch [31,32], but the design process remains challenging and expertise in *de novo* protein design is critical. The RosettaSurf approach allows to rescue protein scaffolds that only present partial structural matches to the epitope structure motif.

As a demonstration of the use of RosettaSurf to transplant the epitope site into an unrelated scaffold, a two-step strategy was employed: 1) side chain grafting of the  $\alpha$ -helical segment onto a small, monomeric protein scaffold (Fig 5A); 2) RosettaSurf design of the remaining antigenic site, including the surface generated by the epitope loop and transition to the helical fragment



**Fig 5. Surface centric design of a viral antigenic site present in RSVF.** A) Design process of site 0-mimicking protein scaffolds. Starting scaffolds are selected from the PDB based on structural alignments with the epitope helix. The surface mimicking designs are generated by grafting the side chains of the helix segment of the epitope onto the scaffold and surface-centric design is employed to optimize the loop region. Before and after design of the surface compared to native site 0. Blue areas indicate high similarity. B) Mimicry of surface geometry of WT scaffold, RSV\_FixBB, and Surf\_03 designs compared to native site 0. C) Representative SPR measurements of Surf\_03 and RSV\_FixBB against site 0-specific antibodies D25 and ADI14496. D) Binding profiles of Surf\_03, a FixBB designed protein, and a helix-only design against a panel of site-specific antibodies with green indicating binding and red cells corresponding to non-binding. A knockout mutant of Surf\_03 and the WT protein binding profiles are listed as reference.

<https://doi.org/10.1371/journal.pcbi.1009178.g005>

(Fig 5A). The grafted helical segment serves as anchor point around which the surface can be optimized to mimic the complete antigenic site.

We identified the NarX histidine kinase receptor (PDB: 3EZI) [46] as promising scaffold. The small, monomeric protein can accommodate the epitope  $\alpha$ -helix ( $C_{\alpha}$  RMSD: 0.4 Å) and offers a sufficiently large surface area around the helix to be optimized with RosettaSurf. We generated 760 designs using the RosettaSurf pipeline and the site 0 surface, as observed in complex with site-specific antibody D25 [47], as a sequence optimization target. The best design was selected based on the highest surface similarity score and was further optimized with six mutations (Surf\_03). First, five point mutations, which were not part of the interface, were introduced to resolve clashes between native scaffold and epitope residues in the  $\alpha$ -helix as well as steric hindrance with residues introduced during surface-centric design. In addition, we mutated a polar residue in the binding pocket to a hydrophobic residue based on the sequence profile of the top 20 surface-designed decoys.

Two additional variants of the same protein scaffold were designed to test whether there were advantages in using RosettaSurf. One design containing only the  $\alpha$ -helix of site 0 (RSV\_helix) and three point mutations that resolve clashes with the side chain grafted epitope helix, similar to Surf\_03. The second design was generated with FixBB (RSV\_FixBB), starting with side-chain grafting of the epitope  $\alpha$ -helix, and allowing to design the same amino acid positions as those of Surf\_03. The RosettaSurf design (Surf\_03) reaches a surface similarity score of 0.45 compared to the antigenic site in the native viral protein RSVF, while the native scaffold scores 0. The helix-grafted base-design scores 0.06 and the RSV\_FixBB design 0.04 in surface resemblance.

To test experimentally our predictions, the designs were expressed and purified, and the binding affinities were measured using a panel of monoclonal antibodies that engage the site 0 epitope (Fig 5B) [31,47,48]. In essence, this panel of monoclonal antibodies (mAbs) was used as conformational probes to assess the surface mimicry presented by the designs. We compared the binding profiles of the three designs, a negative control (Surf\_03 KO) and the WT scaffold using surface plasmon resonance (S5 Fig). All three designs bound to the D25 and ADI19009 mAbs, indicating that these mAbs mostly rely on the helical segment of the epitope. Surf\_03 was the only design recognized by two additional antibodies (ADI14496 and ADI18900), indicating that the higher surface mimicry achieved through RosettaSurf design improves the presentation of the antigenic site and promotes binding of additional site-specific antibodies. The designed immunogen demonstrates RosettaSurf's capabilities to sculpt protein surfaces with a high degree of accuracy which could be of use to introduce functional sites into protein scaffolds by optimizing the molecular surface.

## Discussion

In this work we propose a surface-centric protein design approach and demonstrate its accuracy in several *in silico* benchmarking and experimental design tasks. The protocol can either be used to optimize surface similarity or surface complementarity depending on the design task at hand.

Our ideas to leverage the surface geometrical and chemical features stem from observations that surfaces displaying similar patterns can have similar functional roles (e.g. in PPIs [1,11,14,49]). Technically, our framework is derived from earlier work by Lawrence and Coleman that introduced a fast and efficient way to evaluate shape complementarity as well as work by McCoy et al. that addressed electrostatic complementarity in interfaces. Based on some of these principles, we further implemented a surface similarity score to evaluate both shape and chemistry of surfaces. The resulting surface similarity score is implemented with a

focus on surface-centric protein design pipeline, allowing easy interaction and modification of the algorithm inside Rosetta design protocols.

We showed the high accuracy of the surface score in recovering single amino acid identities by their surface properties and show its benefit when used in conjunction with the Rosetta energy score. Further, we highlight the use of surface similarity score inside a Monte Carlo sampling approach and its performance in the sequence recovery of complete protein interfaces. The incorporation of surface similarity clearly increases sequence recovery rates relative to other scoring schemes.

This approach can be readily applied for the design of experimental combinatorial libraries to reduce library sizes, as the high sequence recovery rates suggest that surface-centric design is especially interesting to design novel proteins that mimic surface patches of other proteins, as it is the case for the design of novel immunogens for vaccine development [30–32].

We demonstrated that surface-centric design could be used to generate targeted libraries to optimize binding specificity of a previously reported IL-2 design [41]. With the ranking of all mutations, we were able to recover amino acid variants that were also obtained by experimental screening through a saturation mutagenesis library. This approach represents a straightforward computational screening method to detect mutations that modulate specificity and affinity in protein-protein interactions. Specifically, this strategy is a fast and accessible alternative to experimental screening techniques.

The provided benchmarks highlight the protocol's design capabilities when operating on static backbones. However, the modelling of protein flexibility remains a challenge for computational methods at large. Due to RosettaSurf's modular implementation within the Rosetta framework, we anticipate that the protocol may perform well when working with conformational flexibility but will be dependent on the quality of the conformational ensembles generated by upstream methods.

Finally, we applied the surface-centric design protocol to engineer novel immunogens to present the surface patch of an antigenic site present in the RSVF protein. We used a structurally complex epitope (site 0 from RSVF), that consists of two structural segments, as an example for structurally challenging sites that remain difficult for computational design approaches. The surface-centric designs show a broader binding profile across a panel of monoclonal antibodies as compared to other design approaches, suggesting that the surface presented is a closer mimic of the native antigenic site.

Possible applications for surface-centric design range from the computational design of highly specific protein-protein interactions, focusing on optimizing the surface complementarity both in shape and chemistry. While the herein presented examples work under the premise of a known protein structure to be optimized for binding, an important related problem is the design of a target surface in the absence of a known backbone template. To our knowledge, identifying backbones with compatible target surfaces remains an open research question, however, we envision two approaches that can readily be combined with RosettaSurf. One simplified approach could utilize a database of protein backbone scaffolds that could be used as initial docking candidates and subsequently optimized with RosettaSurf to improve complementarity towards the target. A second potential approach is to apply a recently developed geometric deep learning framework called MaSIF [14]. Due to its excellent speed, MaSIF allows the comparison of a given surface to all surfaces in the PDB for their complementarity, and this way identifying potential design templates that could be further optimized with RosettaSurf.

In the scope of translational applications, surface similarity enables the computational design of proteins that may recapitulate precise surface features of target surfaces, which could then find applications in immunogen design, where surface similarity allows the design of

immunogens mimicking an antigenic site of interest that could ultimately be used as probes for antibody isolation or vaccine antigens.

Ultimately, we introduce a new conceptual approach in computational protein design where the fine features of molecular surfaces are explicitly optimized. Through this route it will now be possible to explore the tantalizing hypothesis of the existence of a “surface degeneracy code”, which in some cases may allow to represent similar surface patches using very distinct sets of amino acids. Importantly, this capability could also enable the presentation of similar surface patches in proteins with completely different backbone architectures which could again realize many endeavors in functional protein design that are thus far out of reach.

## Materials and methods

### Surface similarity calculation

The described surface similarity ( $\text{Surf}_S$ ) score is composed of shape and electrostatic similarity components. To evaluate surface shape similarity, two protein structures are first aligned and the closest surface points between target and reference are identified. Next, the normal vectors of the surface points are compared by first computing their dot product to obtain the enclosed angle, followed by distance-based scaling that penalizes points that are far apart (Fig 1). Each comparison yields a shape similarity score for the considered pair of points and the similarity of the entire surface results from computing the mean of all pair-wise point comparisons. Since the identification of closest points depends on the starting surface, the surface scores are not identical depending on which structure is the considered the reference and the target and thus the scores are computed in both combinations. A robust surface similarity score is obtained by considering the similarity of the target surface compared to the reference surface and vice versa, effectively averaging the shape similarity to correct for differences during the selection of closest points.

To quantify surface electrostatics, the electrostatic potentials are computed over the continuous electrostatic field and discrete charge values are assigned to every point of the surface using the APBS software [50] (Fig 1). The representation of the target and reference surface point clouds as vectors allows fast computation of Pearson's and Spearman's rank correlations as described by McCoy and colleagues [20]. The resulting correlation coefficients capture the similarity of the individual potential values as well as the overall trends in electrostatic similarity of the two surfaces [18,20].

We combined both, the shape and electrostatic similarity scores, into a single  $\text{Surf}_S$  score to facilitate accurate description of molecular surfaces. To identify a set of optimal weights that describe the relative contributions of the geometric and biochemical features towards the overall properties of the molecular surface, we aimed to derive weights from native protein surfaces. As our goal was to ultimately apply the  $\text{Surf}_S$  score to design novel protein binders, we focused on surfaces arising at the interface of PPIs.

The dataset of protein complexes was compiled from the PDBbind [51], the SAbDab [52], and the Affinity Database versions 1.0 and 2.0 [53] databases, containing crystal structures of transient PPIs. We used RosettaScripts to determine the frequency of each amino acid type across all interaction interfaces. For each amino acid type (excluding cysteines) a total of 140 complexes ( $19 \times 140 = 2,660$  complexes) containing the respective amino acid in the interface region were considered to allow for a balanced dataset with equal numbers of complexes for each amino acid type and no distinctions were made based on rotameric conformations. All possible point mutations to non-native amino acids were generated and their geometric and electrostatic similarity to the native surface measured, resulting in a dataset containing shape and electrostatic properties for each mutation. Logistic regression was applied to identify the

optimal set of weights to combine shape and electrostatic features to optimize for the highest recovery rates of the native amino acids. In total, the dataset contained 2,660 true positive and 47,880 true negative ( $18 \times 19 \times 140 = 47,880$ ) data points. The dataset was split into training and testing subsets, with the testing set containing 20% of the data. Logistic regression was performed using Python's scikit-learn library [54], with the model reaching an accuracy of 98%. The resulting model parameters, i.e. an intercept of -13.79986756 and coefficients 1 and 2 with -13.78594078 and 14.64347448 respectively, provide the weights to combine shape and electrostatic similarity measurements.

## Benchmark datasets

The dataset of the single amino acid recovery benchmark consists of 1,900 protein complexes, with 100 complexes for each amino acid identity except cysteine. All structures used during the various benchmarks as well as the proteins used to generate the Surf<sub>s</sub> score were subjected to constrained energy minimization in their complexed state using Rosetta to adapt them to the Rosetta energy function and the resulting energy optimized decoy served as input for the subsequent benchmark analysis. The protein complexes were randomly selected from the same pool of transient PPIs used for the logistic regression analysis while ensuring that no duplicated complexes were used for the logistic regression training and the single amino acid recovery benchmark. We divided each of these protein complexes into target proteins, where the interface can be mutated, and binders, the proteins that serve as context for the target. The interface of the target protein is defined as residues that are within 7 Å  $C_{\beta}$ -distance of the binder and have the  $C_{\beta}$ -atom pointing towards the binder. This selection ensures that the amino acid side chains are part of the interface and the contribution of the residue to the binding interaction is not solely due to backbone interactions. The full target interface is converted to alanine, effectively removing any side chain memory of the native structure that would restrict the placement of new rotamers and introduce biases towards the native sequence. Cysteine residues are ignored as they can form chemical linkages in the form of disulfide bonds, thus generating a surface that is not attributable to an individual residue.

To test the performance of the surface scoring function we sought to perform a benchmark to evaluate sequence recovery at the interfaces of protein-protein interactions with varying shape complementarity. We assembled a diverse dataset consisting of nine protein complexes capturing different aspects of PPIs and grouped them into three different categories: 1) low-shape complementary interactions which include Enterotoxin G-T-cell receptor complex (PDB: 3MC0) [55], Ribonuclease A in complex with its inhibitor (PDB: 1DFJ) [56] and domain 2 of VEGFR1 in complex with PIGF (PDB: 1RV6) [57]; 2) High-shape complementarity interactions which include the complex between Colicin E9 and IM9 (PDB: 1EMV) [58], Bovine beta-trypsin in complex with CMTI-I (PDB: 1PPE) [59] and PD-L1 in complex with a nanobody (PDB: 5JDS) [60]; 3) antigen-antibody interactions which include HIV-gp120 in complex with CD4-binding site antibody b13 (PDB: 3IDX) [61], RSV epitope-scaffold in complex with Motavizumab (PDB: 4JLR) [28] and the Vaccinia virus D8 protein in complex with the antibody vv138 (PDB: 6B9J) [62]. The shape complementarity of the complexes in the first two categories was assessed by the Rosetta shape-complementarity filter. Protein complexes with a shape complementarity score less than 0.65 were classified as low-complementarity. During the benchmark analysis we distinguish between bound and unbound proteins, where unbound proteins are obtained by removing the protein binder from the holo-crystal structure.

Data analysis was performed with the help of the rstoolbox Python library [63].



## Computational design of proteins mimicking the antigenic site 0 in RSV

The structure of antigenic site 0 was extracted from the crystal structure of prefusion stabilized RSVF in complex with antibody D25 (PDB: 4JHW) [47]. The D25 epitope consists of an irregular  $\alpha$ -helix (residues 196–209) and a 10-residue loop (residues 61–70). To identify putative scaffolds, we performed a structural search based on the irregular  $\alpha$ -helix against 55,574 monomeric, helix-containing crystal structures from the Protein Data Bank (PDB, from September 2015) using Rosetta's MotifGraft algorithm [64]. Matches were filtered at a backbone RMSD threshold below 0.55 Å and less than ten atomic clashes at the interface, resulting in 13 scaffold candidates. After visual inspection, we selected the NarX histidine kinase receptor (PDB: 3EZI) [46] as scaffold, a 107-residue long, monomeric protein that aligned to the epitope helix with a  $C_{\alpha}$  RMSD of 0.4 Å and provided additional surface area to mimic the entire antigenic site.

In a first step, we transplanted the side chains of the epitope helix contacting the D25 antibody onto the selected scaffold using MotifGraft, followed by the introduction of three mutations on the protein scaffold that were not part of the interface to resolve steric clashes with the transferred side chains, resulting in the design RSV\_helix. Subsequently, we performed surface-centric design on 19 residues using the RosettaSurf pipeline to increase surface mimicry of the site 0 around the epitope helix, resulting in 760 design decoys. Surface-centric design was performed in the presence of the D25 antibody and site 0 of RSVF served as a reference to which our designed surface was optimized. We selected the decoy with the highest surface mimicry score, named surf\_01, and optimized the design in additional steps. To allow accurate display of the designed surface, we introduced two point mutants in the scaffold adjacent to the optimized surface patch to avoid steric hindrance (surf\_02). Lastly, after comparing the similarity of surf\_02 and the native antigenic site, we identified a lysine residue at position 18 in our design with suboptimal mimicry. Evaluating the sequence profile of the 20 best scoring decoys sorted by surface mimicry revealed a strong preference for leucine at this position and the residue was incorporated into a new design (Surf\_03). Finally, we designed a version of RSV\_helix that used Rosetta's fixed-backbone design (FixBB) to serve as comparison to our surface-optimized designs. We designed 1'000 decoys in the presence of D25 antibody, allowing mutations to occur at the same 19 residues as was the case for RosettaSurf design. All designs converged to an identical sequence which was selected as design RSV\_FixBB. Based on Surf\_03 we designed a knockout mutant (Surf\_03\_KO) with a N74Y mutation in the epitope helix.

## Protein expression and purification

**Designs.** Genes for all designs were purchased as DNA fragments from Twist Bioscience, and cloned into pET11 vectors, containing a N-terminal MBP-tag and His-tag as well as a TEV cleavage site, for bacterial expression. Plasmids were transformed into *E. coli* BL21 (DE3 pLysS) (Merck, #69451–3) and grown overnight in LB media at 37°C. Pre-cultures were diluted 1:50 and inoculated to an  $OD_{600}$  of 0.6 in terrific broth (Conalab, #PRO1246.05) at 37°C and expression was induced by the addition of 1 mM isopropyl- $\beta$ -D-thiogalactoside (IPTG). Cultures were harvested after 18–20 hours at 20°C. Pellets were resuspended in lysis buffer (50 mM Tris, pH 7.5, 500 mM NaCl, 5% Glycerol, 1 mg/ml lysozyme, 1 mM PMSF, and 1  $\mu$ g/ml DNase) and sonicated on ice for a total of 12 minutes, in intervals of 15 s sonication followed by 45 s pause. The lysates were cleared by centrifugation (48'384 g, 20 min) and purified using a His-Trap FF column on an Äkta pure system (GE Healthcare), followed by size exclusion on a HiLoad 16/600 Superdex 75 column (GE Healthcare) in phosphate-buffered saline (PBS). Protein concentrations were determined by measuring the absorbance at 280 nm

on a Nanodrop (Thermo Scientific). The designed proteins were concentrated by centrifugation (Millipore, #UFC900324) to 1 mg/ml, snap frozen in liquid nitrogen, and stored at  $-80^{\circ}\text{C}$ .

**Antibody variable fragments (Fabs).** For Fab expression, heavy and light chain DNA sequences were purchased from Twist Biosciences and cloned separately into the pHLSec mammalian expression vector (Addgene, #99845) using AgeI and XhoI restriction sites. Expression plasmids were premixed in a 1:1 stoichiometric ratio, co-transfected into HEK293-F cells, and cultured in FreeStyle medium (Gibco, #12338018). Supernatants were harvested after 1 week by centrifugation and purified using a kappa-select column (GE Healthcare). Elution of bound proteins was conducted using 0.1 M glycine buffer (pH 2.7), and eluates were immediately neutralized by the addition of 1 M Tris ethylamine (pH 9), followed by buffer exchange to PBS (pH 7.4).

### Binding affinity determination by surface plasmon resonance (SPR)

SPR measurements were performed on a Biacore 8K (GE Healthcare) with HBS-EP+ as running buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.005% v/v Surfactant P20, GE Healthcare) at room temperature. Approximately 700 response units (RU) of Fabs were immobilized on a CM5 sensor chip (GE Healthcare) via amine coupling, and designed monomeric proteins were injected as analyte in two-fold serial dilutions. The flow rate was 30  $\mu\text{l}/\text{min}$  with 120 s of contact time followed by 400 s dissociation time. After each injection, surface was regenerated using 0.1 M glycine at pH 3.5. Data were fitted using 1:1 Langmuir binding model within the Biacore 8K analysis software (GE Healthcare #29310604).

### Supporting information

**S1 Fig. Schematic overview of the surface-centric design process.** A surface patch is selected on the target protein that will be subjected to mutations for improving surface features. A reference surface is specified and will be used during the design process to guide the introduced mutations. During sequence design, rotamers are sampled in the selected interfaces of the target protein and for each substitution the surface is compared to the reference surface. If mutations improve the surface score, the changes are accepted. During iterative sampling steps of the selected surface patch, the overall surface can be improved.

(TIF)

**S2 Fig. Recovery success of individual amino acid types.** Case study of the recovery of arginine in difficult and easily recoverable benchmark cases. A) Recovery of the exposed arginine residue is unsuccessful in the unbound test case as a non-native rotamer is placed in the structure. The addition of the binding partner limits the accessible rotameric space and allows successful recovery of the amino acid. B) Successful recovery of arginine independent of the presence or absence of the binder as the native rotamer conformation is less exposed. Overall side chain configurations placed in the bound benchmark cases are closer to the native rotamer (mean full-atom RMSD of  $\sim 0.4$  Å) as compared to the unbound benchmark cases (mean full-atom RMSD of  $\sim 2.3$  Å).

(TIF)

**S3 Fig. Impact of small changes in rotamers on the surface score.** A small rotamer change shown for lysine results in only a 0.8 Å all-atom RMSD change (left). When evaluating both rotamers in terms of surface similarity, the  $\text{Surf}_S$  score can discriminate the local changes. The shape similarity score changes by 0.2 and the electrostatic similarity score by 0.5 units, resulting in an overall  $\text{Surf}_S$  score of 0.989 when comparing both rotamers. The differences are specific for the altered region as shown for shape similarity (right) and outline the modified atom

positions.  
(TIF)

**S4 Fig. Sequence recovery of protein interfaces grouped by different metrics.** A) Sequence recovery of interfaces from protein complexes grouped by the amount of flexible amino acid side chains. The plot on the left site shows the overall sequence recovery for all complexes combined under the three different design approaches, i.e. FixBB, RosettaSurf, and RosettaSurf-site. The *Flexible* plot represents the sequence recovery rate of protein complexes whose interfaces consists at least of 40% amino acids with flexible side chains (PDB IDs: 1DFJ, 1EMV, 1PPE, 3MC0). The *Static* plot contains complexes with less than 40% flexible amino acids in the interface (PDB IDs: 1RV6, 3IDX, 4JLR, 5JDS, 6B9J). Amino acids were considered flexible/static based on Scouras and Daggett's observations on rotamer dynamics. Accordingly, we assigned R, N, Q, E, H, K, M, and W as flexible and D, I, L, F, P, S, T, Y, V, A, and G as static amino acids. Spearman rank's correlation analysis shows a weak correlation of sequence recovery rate and rotamer flexibility ( $\rho = -0.12$ ,  $p = 5.58e-163$ ), indicating that rotamers with higher conformational flexibility are more difficult to recover. B) Sequence recovery of protein interfaces grouped by the median distance of closest surface points into low (median distance  $d < 0.5$ ; PDB IDs: 1EMV, 5JDS, 6B9J), medium (median distance  $0.5 \leq d < 0.6$ ; PDB IDs: 1PPE, 3IDX, 4JLR), and high (median distance  $0.6 \leq d$ ; PDB IDs: 1DFJ, 1RV6, 3MC0) groups. C) Sequence recovery of protein interfaces grouped by interface area into low ( $A < 800 \text{ \AA}^2$ ; PDB IDs: 3IDX, 3MC0, 5JDS), medium ( $800 \text{ \AA}^2 \leq A < 1000 \text{ \AA}^2$ ; PDB IDs: 1EMV, 1RV6, 4JLR), and high ( $1000 \text{ \AA}^2 \leq A$ ; PDB IDs: 1DFJ, 1PPE, 6B9J) groups. D) Sequence recovery of protein interfaces grouped by hydrophobic interface area into low ( $A < 600 \text{ \AA}^2$ ; PDB IDs: 1DFJ, 3MC0, 6B9J), medium ( $600 \text{ \AA}^2 \leq A < 810 \text{ \AA}^2$ ; PDB IDs: 3IDX, 4JLR, 5JDS), and high ( $810 \text{ \AA}^2 \leq A$ ; PDB IDs: 1EMV, 1PPE, 1RV6) groups. For all benchmark cases, the sequence recovery is reported in the absence (unbound) and presence (bound) of the binder.  
(TIF)

**S5 Fig. Surface plasmon resonance measurements of designed proteins against five site 0 targeting antibodies.** The surface designed protein (Surf\_03) demonstrates the broadest binding reactivity, showing measurable interactions with antibodies D25, ADI14496, ADI18900, and ADI19009. In contrast, a design containing solely the helical motif (RSV\_helix) of the site 0 epitope interacts only with D25 and ADI19009. Similar results are obtained for a Rosetta-designed variant (RSV\_FixBB), binding to D25 and ADI19009. Controls in the form of a KO mutant of design Surf\_03 (Surf\_03 KO) and the wild type protein (WT) show no measurable binding, indicating that the observed interactions are specific for the designed epitope sites.  
(TIF)

## Acknowledgments

We would like to thank Pablo Gainza, Fabian Sesterhenn, Che Yang, and Sarah Wehrle for helpful discussions and support regarding *in vitro* experiments. We thank Sailan Shui, Freyr Sverrisson, and Arne Schneuing for comments on the manuscript. We also thank Luki Goldschmidt and Brian Coventry for implementing the shape and electrostatic complementarity evaluation within Rosetta, respectively. Additionally, we would like to acknowledge the high-performance computing facility (SCITAS) for their technical support. We would also like to acknowledge the Swiss National Supercomputing Centre (CSCS) for their support in computing time.

## Author Contributions

**Conceptualization:** Andreas Scheck, Bruno E. Correia.

**Data curation:** Andreas Scheck.

**Formal analysis:** Andreas Scheck, Stéphane Rosset, Michaël Defferrard, Andreas Loukas.

**Funding acquisition:** Pierre Vandergheynst, Bruno E. Correia.

**Methodology:** Michaël Defferrard, Andreas Loukas.

**Project administration:** Andreas Scheck, Bruno E. Correia.

**Software:** Andreas Scheck, Jaume Bonet.

**Supervision:** Pierre Vandergheynst, Bruno E. Correia.

**Writing – original draft:** Andreas Scheck, Bruno E. Correia.

**Writing – review & editing:** Andreas Scheck, Stéphane Rosset, Michaël Defferrard, Andreas Loukas, Jaume Bonet, Pierre Vandergheynst, Bruno E. Correia.

## References

1. Shulman-Peleg A, Nussinov R, Wolfson HJ. Recognition of functional sites in protein structures. *Journal of molecular biology*. 2004; 339(3):607–33. <https://doi.org/10.1016/j.jmb.2004.04.012> PMID: 15147845
2. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. *Journal of molecular biology*. 1971; 55(3):379–IN4. [https://doi.org/10.1016/0022-2836\(71\)90324-x](https://doi.org/10.1016/0022-2836(71)90324-x) PMID: 5551392
3. Connolly ML. Analytical molecular surface calculation. *Journal of applied crystallography*. 1983; 16(5):548–58.
4. Connolly ML. Solvent-accessible surfaces of proteins and nucleic acids. *Science*. 1983; 221(4612):709–13. <https://doi.org/10.1126/science.6879170> PMID: 6879170
5. Richards FM. Areas, Volumes, Packing, and Protein Structure. *Annu Rev Biophys Bioeng*. 1977 Jun; 6(1):151–76.
6. Ryu J, Park R, Kim D-S. Molecular surfaces on proteins via beta shapes. *Computer-Aided Design*. 2007; 39(12):1042–57.
7. Greer J, Bush BL. Macromolecular shape and surface maps by solvent exclusion. *Proceedings of the National Academy of Sciences*. 1978 Jan 1; 75(1):303–7. <https://doi.org/10.1073/pnas.75.1.303> PMID: 272646
8. Kihara D, Sael L, Chikhi R, Esquivel-Rodriguez J. Molecular surface representation using 3D Zernike descriptors for protein shape comparison and docking. *Current Protein and Peptide Science*. 2011; 12(6):520–30. <https://doi.org/10.2174/138920311796957612> PMID: 21787306
9. Schmitt S, Kuhn D, Klebe G. A New Method to Detect Related Function Among Proteins Independent of Sequence and Fold Homology. *Journal of Molecular Biology*. 2002 Oct 18; 323(2):387–406. [https://doi.org/10.1016/S0022-2836\(02\)00811-2](https://doi.org/10.1016/S0022-2836(02)00811-2) PMID: 12381328
10. Kinoshita K, Nakamura H. Identification of protein biochemical functions by similarity search using the molecular surface database eF-site. *Protein Science*. 2003; 12(8):1589–95. <https://doi.org/10.1110/ps.0368703> PMID: 12876308
11. Yin S, Proctor EA, Lugovskoy AA, Dokholyan NV. Fast screening of protein surfaces using geometric invariant fingerprints. *Proceedings of the National Academy of Sciences*. 2009; 106(39):16622–6. <https://doi.org/10.1073/pnas.0906146106> PMID: 19805347
12. Daberdaku S, Ferrari C. Fast Computation of High-resolution Solvent Excluded Protein Surface with OpenMP. In: 2018 International Conference on High Performance Computing Simulation (HPCS). 2018. p. 789–96.
13. Natarajan V, Koehl P, Wang Y, Hamann B. Visual analysis of biomolecular surfaces. In: *Visualization in Medicine and Life Sciences*. Springer; 2008. p. 237–55.
14. Gainza P, Sverrisson F, Monti F, Rodola E, Boscaini D, Bronstein M, et al. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nature Methods*. 2020; 17(2):184–92. <https://doi.org/10.1038/s41592-019-0666-6> PMID: 31819266

15. Sverrisson F, Feydy J, Correia BE, Bronstein MM. Fast End-to-End Learning on Protein Surfaces. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021. p. 15272–81.
16. Morris RJ, Najmanovich RJ, Kahraman A, Thornton JM. Real spherical harmonic expansion coefficients as 3D shape descriptors for protein binding pocket and ligand comparisons. *Bioinformatics*. 2005; 21(10):2347–55. <https://doi.org/10.1093/bioinformatics/bti337> PMID: 15728116
17. Sharp KA, Honig B. Electrostatic interactions in macromolecules: theory and applications. *Annual review of biophysics and biophysical chemistry*. 1990; 19(1):301–32. <https://doi.org/10.1146/annurev.bb.19.060190.001505> PMID: 2194479
18. Chau P-L, Dean PM. Electrostatic complementarity between proteins and ligands. 1. Charge disposition, dielectric and interface effects. *J Computer-Aided Mol Des*. 1994 Oct 1; 8(5):513–25. <https://doi.org/10.1007/BF00123663> PMID: 7876898
19. Lawrence MC, Colman PM. Shape complementarity at protein/protein interfaces. Elsevier; 1993.
20. McCoy AJ, Epa VC, Colman PM. Electrostatic complementarity at protein/protein interfaces. *Journal of molecular biology*. 1997; 268(2):570–84. <https://doi.org/10.1006/jmbi.1997.0987> PMID: 9159491
21. Venkatraman V, Sael L, Kihara D. Potential for protein surface shape analysis using spherical harmonics and 3D Zernike descriptors. *Cell biochemistry and biophysics*. 2009; 54(1–3):23–32. <https://doi.org/10.1007/s12013-009-9051-x> PMID: 19521674
22. Huang P-S, Oberdorfer G, Xu C, Pei XY, Nannenga BL, Rogers JM, et al. High thermodynamic stability of parametrically designed helical bundles. *science*. 2014; 346(6208):481–5. <https://doi.org/10.1126/science.1257481> PMID: 25342806
23. Joh NH, Wang T, Bhate MP, Acharya R, Wu Y, Grabe M, et al. De novo design of a transmembrane Zn<sup>2+</sup>-transporting four-helix bundle. *Science*. 2014; 346(6216):1520–4. <https://doi.org/10.1126/science.1261172> PMID: 25525248
24. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a novel globular protein fold with atomic-level accuracy. *science*. 2003; 302(5649):1364–8. <https://doi.org/10.1126/science.1089427> PMID: 14631033
25. Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, Clair JLS, et al. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science*. 2010; 329(5989):309–13. <https://doi.org/10.1126/science.1190239> PMID: 20647463
26. Fleishman SJ, Whitehead TA, Ekiert DC, Dreyfus C, Corn JE, Strauch E-M, et al. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science*. 2011; 332(6031):816–21. <https://doi.org/10.1126/science.1202617> PMID: 21566186
27. Tinberg CE, Khare SD, Dou J, Doyle L, Nelson JW, Schena A, et al. Computational design of ligand-binding proteins with high affinity and selectivity. *Nature*. 2013; 501(7466):212–6. <https://doi.org/10.1038/nature12443> PMID: 24005320
28. Correia BE, Bates JT, Loomis RJ, Baneyx G, Carrico C, Jardine JG, et al. Proof of principle for epitope-focused vaccine design. *Nature*. 2014; 507(7491):201–6. <https://doi.org/10.1038/nature12966> PMID: 24499818
29. Bonet J, Wehrle S, Schriever K, Yang C, Billet A, Sesterhenn F, et al. Rosetta FunFolDes—A general framework for the computational design of functional proteins. *PLoS computational biology*. 2018; 14(11):e1006623. <https://doi.org/10.1371/journal.pcbi.1006623> PMID: 30452434
30. Sesterhenn F, Galloux M, Vollers SS, Csepregi L, Yang C, Descamps D, et al. Boosting subdominant neutralizing antibody responses with a computationally designed epitope-focused immunogen. *PLoS biology*. 2019; 17(2):e3000164. <https://doi.org/10.1371/journal.pbio.3000164> PMID: 30789898
31. Sesterhenn F, Yang C, Bonet J, Cramer JT, Wen X, Wang Y, et al. De novo protein design enables the precise induction of RSV-neutralizing antibodies. *Science*. 2020; 368(6492). <https://doi.org/10.1126/science.aay5051> PMID: 32409444
32. Yang C, Sesterhenn F, Bonet J, van Aalen EA, Scheller L, Abriata LA, et al. Bottom-up de novo design of functional proteins with complex structural features. *Nature Chemical Biology*. 2021; 1–9. <https://doi.org/10.1038/s41589-020-00714-1> PMID: 33328655
33. Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods in enzymology*. 2011; 487:545–74. <https://doi.org/10.1016/B978-0-12-381270-4.00019-6> PMID: 21187238
34. Fleishman SJ, Leaver-Fay A, Corn JE, Strauch E-M, Khare SD, Koga N, et al. RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. *PloS one*. 2011; 6(6):e20161. <https://doi.org/10.1371/journal.pone.0020161> PMID: 21731610

35. Alford RF, Leaver-Fay A, Jeliaskov JR, O'Meara MJ, DiMaio FP, Park H, et al. The Rosetta all-atom energy function for macromolecular modeling and design. *Journal of chemical theory and computation*. 2017; 13(6):3031–48. <https://doi.org/10.1021/acs.jctc.7b00125> PMID: 28430426
36. Kiyoshi M, Caaveiro JMM, Miura E, Nagatoishi S, Nakakido M, Soga S, et al. Affinity Improvement of a Therapeutic Antibody by Structure-Based Computational Design: Generation of Electrostatic Interactions in the Transition State Stabilizes the Antibody-Antigen Complex. *PLOS ONE*. 2014 Jan 27; 9(1): e87099. <https://doi.org/10.1371/journal.pone.0087099> PMID: 24475232
37. Fischer A, Seitz T, Lochner A, Sterner R, Merkl R, Bocola M. A Fast and Precise Approach for Computational Saturation Mutagenesis and its Experimental Validation by Using an Artificial ( $\beta\alpha$ )8-Barrel Protein. *ChemBioChem*. 2011; 12(10):1544–50. <https://doi.org/10.1002/cbic.201100051> PMID: 21626637
38. Szymczyk P, Szymańska G, Lipert A, Weremczuk-Jeżyna I, Kochan E. Computer-Aided Saturation Mutagenesis of *Arabidopsis thaliana* Ent-Copalyl Diphosphate Synthase. *Interdiscip Sci Comput Life Sci*. 2020 Mar 1; 12(1):32–43. <https://doi.org/10.1007/s12539-019-00342-x> PMID: 31309397
39. Vedithi SC, Rodrigues CHM, Portelli S, Skwark MJ, Das M, Ascher DB, et al. Computational saturation mutagenesis to predict structural consequences of systematic mutations in the beta subunit of RNA polymerase in *Mycobacterium leprae*. *Computational and Structural Biotechnology Journal*. 2020 Jan 1; 18:271–86. <https://doi.org/10.1016/j.csbj.2020.01.002> PMID: 32042379
40. Öhlknecht C, Katz S, Kröß C, Sprenger B, Engele P, Schneider R, et al. Efficient In Silico Saturation Mutagenesis of a Member of the Caspase Protease Family. *J Chem Inf Model*. 2021 Mar 22; 61(3):1193–203. <https://doi.org/10.1021/acs.jcim.0c01216> PMID: 33570387
41. Silva D-A, Yu S, Ulge UY, Spangler JB, Jude KM, Labão-Almeida C, et al. De novo design of potent and selective mimics of IL-2 and IL-15. *Nature*. 2019; 565(7738):186–91. <https://doi.org/10.1038/s41586-018-0830-7> PMID: 30626941
42. Procko E, Berguig GY, Shen BW, Song Y, Frayo S, Convertine AJ, et al. A computationally designed inhibitor of an Epstein-Barr viral Bcl-2 protein induces apoptosis in infected cells. *Cell*. 2014; 157(7):1644–56. <https://doi.org/10.1016/j.cell.2014.04.034> PMID: 24949974
43. Whitehead TA, Chevalier A, Song Y, Dreyfus C, Fleishman SJ, De Mattos C, et al. Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat Biotechnol*. 2012 Jun; 30(6):543–8. <https://doi.org/10.1038/nbt.2214> PMID: 22634563
44. Yan Y, Huang S-Y. Protein-protein docking with improved shape complementarity. In: *International Conference on Intelligent Computing*. Springer; 2018. p. 600–5.
45. Correia BE, Ban Y-EA, Holmes MA, Xu H, Ellingson K, Kraft Z, et al. Computational Design of Epitope-Scaffolds Allows Induction of Antibodies Specific for a Poorly Immunogenic HIV Vaccine Epitope. *Structure*. 2010 Sep 8; 18(9):1116–26. <https://doi.org/10.1016/j.str.2010.06.010> PMID: 20826338
46. Cheung J, Hendrickson WA. Structural analysis of ligand stimulation of the histidine kinase NarX. *Structure*. 2009; 17(2):190–201. <https://doi.org/10.1016/j.str.2008.12.013> PMID: 19217390
47. McLellan JS, Chen M, Leung S, Graepel KW, Du X, Yang Y, et al. Structure of RSV Fusion Glycoprotein Trimer Bound to a Prefusion-Specific Neutralizing Antibody. *Science*. 2013 May 31; 340(6136):1113–7. <https://doi.org/10.1126/science.1234914> PMID: 23618766
48. Gilman MSA, Castellanos CA, Chen M, Ngwuta JO, Goodwin E, Moin SM, et al. Rapid profiling of RSV antibody repertoires from the memory B cells of naturally infected adult donors. *Science Immunology* [Internet]. 2016 Dec 9 [cited 2021 May 5]; 1(6). Available from: <https://immunology.sciencemag.org/content/1/6/eaaj1879> <https://doi.org/10.1126/sciimmunol.aaj1879> PMID: 28111638
49. Keskin O, Nussinov R. Similar Binding Sites and Different Partners: Implications to Shared Proteins in Cellular Pathways. *Structure*. 2007 Mar 1; 15(3):341–54. <https://doi.org/10.1016/j.str.2007.01.007> PMID: 17355869
50. Jurrus E, Engel D, Star K, Monson K, Brandi J, Felberg LE, et al. Improvements to the APBS biomolecular solvation software suite. *Protein Science*. 2018; 27(1):112–28. <https://doi.org/10.1002/pro.3280> PMID: 28836357
51. Wang R, Fang X, Lu Y, Wang S. The PDBbind Database: Collection of Binding Affinities for Protein–Ligand Complexes with Known Three-Dimensional Structures. *J Med Chem*. 2004 Jun 1; 47(12):2977–80. <https://doi.org/10.1021/jm030580l> PMID: 15163179
52. Dunbar J, Krawczyk K, Leem J, Baker T, Fuchs A, Georges G, et al. SABDab: the structural antibody database. *Nucleic Acids Research*. 2014 Jan 1; 42(D1):D1140–6. <https://doi.org/10.1093/nar/gkt1043> PMID: 24214988
53. Vreven T, Moal IH, Vangone A, Pierce BG, Kastriitis PL, Torchala M, et al. Updates to the Integrated Protein–Protein Interaction Benchmarks: Docking Benchmark Version 5 and Affinity Benchmark Version 2. *Journal of Molecular Biology*. 2015 Sep 25; 427(19):3031–41. <https://doi.org/10.1016/j.jmb.2015.07.016> PMID: 26231283

54. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *MACHINE LEARNING IN PYTHON*.: 6.
55. Fernández MM, Cho S, De Marzi MC, Kerzic MC, Robinson H, Mariuzza RA, et al. Crystal structure of staphylococcal enterotoxin G (SEG) in complex with a mouse T-cell receptor  $\beta$  chain. *Journal of Biological Chemistry*. 2011; 286(2):1189–95. <https://doi.org/10.1074/jbc.M110.142471> PMID: 21059660
56. Kobe B, Deisenhofer J. A structural basis of the interactions between leucine-rich repeats and protein ligands. *Nature*. 1995; 374(6518):183–6. <https://doi.org/10.1038/374183a0> PMID: 7877692
57. Christinger HW, Fuh G, de Vos AM, Wiesmann C. The crystal structure of placental growth factor in complex with domain 2 of vascular endothelial growth factor receptor-1. *Journal of Biological Chemistry*. 2004; 279(11):10382–8. <https://doi.org/10.1074/jbc.M313237200> PMID: 14684734
58. Kühlmann UC, Pommer AJ, Moore GR, James R, Kleanthous C. Specificity in protein-protein interactions: the structural basis for dual recognition in endonuclease colicin-immunity protein complexes. *Journal of molecular biology*. 2000; 301(5):1163–78. <https://doi.org/10.1006/jmbi.2000.3945> PMID: 10966813
59. Bode W, Greyling HJ, Huber R, Otlewski J, Wilusz T. The refined 2.0 Å X-ray crystal structure of the complex formed between bovine  $\beta$ -trypsin and CMTI-I, a trypsin inhibitor from squash seeds (*Cucurbita maxima*) Topological similarity of the squash seed inhibitors with the carboxypeptidase A inhibitor from potatoes. *FEBS letters*. 1989; 242(2):285–92. [https://doi.org/10.1016/0014-5793\(89\)80486-7](https://doi.org/10.1016/0014-5793(89)80486-7) PMID: 2914611
60. Zhang F, Wei H, Wang X, Bai Y, Wang P, Wu J, et al. Structural basis of a novel PD-L1 nanobody for immune checkpoint blockade. *Cell discovery*. 2017; 3(1):1–12. <https://doi.org/10.1038/celldisc.2017.4> PMID: 28280600
61. Chen L, Do Kwon Y, Zhou T, Wu X, O'Dell S, Cavacini L, et al. Structural basis of immune evasion at the site of CD4 attachment on HIV-1 gp120. *Science*. 2009; 326(5956):1123–7. <https://doi.org/10.1126/science.1175868> PMID: 19965434
62. Matho MH, Schlossman A, Gilchuk IM, Miller G, Mikulski Z, Hupfer M, et al. Structure–function characterization of three human antibodies targeting the vaccinia virus adhesion molecule D8. *Journal of Biological Chemistry*. 2018; 293(1):390–401. <https://doi.org/10.1074/jbc.M117.814541> PMID: 29123031
63. Bonet J, Hartevelde Z, Sesterhenn F, Scheck A, Correia BE. rstoolbox—a Python library for large-scale analysis of computational protein design data and structural bioinformatics. *BMC Bioinformatics*. 2019 Dec; 20(1):240. <https://doi.org/10.1186/s12859-019-2796-3> PMID: 31092198
64. Silva D-A, Correia BE, Procko E. Motif-Driven Design of Protein–Protein Interfaces. In: Stoddard BL, editor. *Computational Design of Ligand Binding Proteins* [Internet]. New York, NY: Springer New York; 2016 [cited 2021 Feb 22]. p. 285–304. (*Methods in Molecular Biology*; vol. 1414). Available from: [http://link.springer.com/10.1007/978-1-4939-3569-7\\_17](http://link.springer.com/10.1007/978-1-4939-3569-7_17)