

Article

An Evolutionary Conservation and Druggability Analysis of Enzymes Belonging to the Bacterial Shikimate Pathway

Rok Frlan 

The Department of Pharmaceutical Chemistry, Faculty of Pharmacy, University of Ljubljana, 1000 Ljubljana, Slovenia; rok.frlan@ffa.uni-lj.si; Tel.: +386-1-4769-674

Abstract: Enzymes belonging to the shikimate pathway have long been considered promising targets for antibacterial drugs because they have no counterpart in mammals and are essential for bacterial growth and virulence. However, despite decades of research, there are currently no clinically relevant antibacterial drugs targeting any of these enzymes, and there are legitimate concerns about whether they are sufficiently druggable, i.e., whether they can be adequately modulated by small and potent drug-like molecules. In the present work, *in silico* analyses combining evolutionary conservation and druggability are performed to determine whether these enzymes are candidates for broad-spectrum antibacterial therapy. The results presented here indicate that the substrate-binding sites of most enzymes in this pathway are suitable drug targets because of their reasonable conservation and druggability scores. An exception was the substrate-binding site of 3-deoxy-*D*-arabino-heptulosonate-7-phosphate synthase, which was found to be undruggable because of its high content of charged residues and extremely high overall polarity. Although the presented study was designed from the perspective of broad-spectrum antibacterial drug development, this workflow can be readily applied to any antimicrobial target analysis, whether narrow- or broad-spectrum. Moreover, this research also contributes to a deeper understanding of these enzymes and provides valuable insights into their properties.



Citation: Frlan, R. An Evolutionary Conservation and Druggability Analysis of Enzymes Belonging to the Bacterial Shikimate Pathway. *Antibiotics* **2022**, *11*, 675. <https://doi.org/10.3390/antibiotics11050675>

Academic Editor: Carlos M. Franco

Received: 20 April 2022

Accepted: 13 May 2022

Published: 17 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: shikimate pathway; druggability; ligandability; antibacterial; inhibitors; broad-spectrum

1. Introduction

Although antibiotics have saved many lives in the past, their overuse and misuse have led to an increased economic burden as well as increases in morbidity and mortality in patients infected with formerly common infections that are not treatable with existing treatment options [1]. There are many promising strategies in the fight against bacterial pathogens that have evolved over the past decades, but the one that is most frequently mentioned in the scientific literature is targeting an essential metabolic pathway that is not exploited by existing drugs [2]. In this context, the enzymes of the shikimate pathway may be a possible solution because they have no counterpart in mammals and are essential for growth and virulence, as shown by several knock-out studies [3–9]. Importantly, this pathway was discovered more than 70 years ago and has been studied in detail since then [10–14].

An overview of literature data suggests that the interest in these enzymes as antibacterial targets has increased over the years (Supplementary Materials, Figure S1). However, these enzymes have been largely ignored by the pharmaceutical industry, which to date has been unable to produce a clinically relevant drug against any of them [15]. Although many inhibitors have been published in the scientific literature, most of which are substrate analogs (Supplementary Materials, Figure S2), the only widely used inhibitor is the herbicide glyphosate, which was discovered nearly 50 years ago and acts via the specific inhibition of plant and malarial EPSP synthase (EPSPS) [16–19].

The shikimate metabolic pathway (Supplementary Materials, Figure S3) is found in plants, algae, bacteria, fungi, and apicomplexan parasites. It involves seven enzymes that catalyze a series of reactions to form chorismate, a precursor of aromatic amino acids, iron-scavenging siderophores, folic acid, vitamin K, and ubiquinone [10–14]. The extreme importance of the synthesis of aromatic amino acids in the bacterial life cycle and their structural as well as kinetic properties have been reviewed in detail by many authors to whom we refer the reader [20–29].

In the absence of clinically relevant antibacterial drugs against any of these enzymes, there is a legitimate concern that these targets are not sufficiently ligandable and/or druggable. Druggability is defined as the ability of a target to be modulated by potent, small “drug-like” molecules. Drugs that bind to a target that is more druggable are likely to have higher ligand efficiencies, require lower doses and exposures, and have a lower risk of failure during the developmental stages of the drug. To bind drug-like molecules, a binding site must also be ligandable, i.e., it must be able to bind inhibitors with sufficient affinity and be complementary or compatible with their physicochemical properties [30].

There are many strategies we can use to assess the ligandability and druggability of targets [31]. One of the most rigorous is the evaluation of existing inhibitors or hit rates of HTS campaigns. It was first introduced in 2006 by Macarron, who published a retrospective analysis of HTS results at GlaxoSmithKline (GSK) [32]. Although many other metrics have subsequently been reported, their precision depends heavily on the number of compounds present [33–35]. For example, Vukovic and Huggins [36] described a reward-based metric for quantifying target ligandability for which BindingDB [37,38] reports > 100 Ki values. Since the number of submicromolar inhibitors is insufficient to make a reliable prediction, such an analysis could not be performed with sufficient reliability for most enzymes of the shikimate pathway (Supplementary Materials, Figure S1).

For proteins with 3D structures that are known or predicted using AI systems such as AlphaFold, several computational methods can be used that provide medicinal chemists with important guidelines for prioritizing targets before they perform costly high-throughput screening [30,36]. Over the years, many crystal structures of all seven enzymes of the shikimate pathway have become available in both apo and liganded forms, making such in silico studies feasible. Although numerous software and online platforms have been developed for this purpose, including FTMap [39], DOGSiteScorer [40], SiteMap [41], Fpocket [42], Sitehound [43], etc., few of these have been used for the analysis of shikimate pathway enzymes. To the best of our knowledge, there are only two studies in which the FTMap server was used to identify hotspots of DHQS from *S. aureus* [44] and SDH from *H. pylori* [45].

However, even if a particular binding site proves to be druggable, it may not be a suitable target for antibacterial drug development because its evolutionary conservation is not high enough. The extent to which the amino acids that constitute a particular target are evolutionarily conserved correlates strongly with its structural and functional importance; it is also a predictor of the likelihood that a potential target will be susceptible to mutation and the development of resistance [46]. Evaluations of evolutionary conservation have already been performed, to some extent, for the enzymes of the shikimate pathway, mostly using multiple sequence alignments (MSA) of sequences from different kingdoms for each enzyme of the pathway [20,47–64].

Evolutionary conservation and druggability analyses are therefore crucial for the validation of antibacterial targets, and it is surprising that the combination of these two methods has not yet become a standard target validation procedure in the literature. Such procedures are also rare in other antimicrobial fields, for example, in the antiviral field. To our knowledge, there is only one study by Kukol and Hughes in which both features were considered simultaneously in a large-scale analysis of the influenza A virus nucleoprotein sequence [65]. There are also a few reports in which the identification of hotspots or binding sites was performed in parallel with the calculation of their evolutionary conservation [66–68].

The aim of this study is to perform a combined comprehensive *in silico* analysis of evolutionary conservation and druggability for the enzymes of the shikimate pathway in pathogenic bacteria. In this context, we intend to identify hotspots with FTMap and binding pockets with SiteMap and FTSite in a representative set of crystal structures of each enzyme of the pathway. Since binding regions are not strictly limited to catalytic sites, the presence of other regions, such as allosteric sites, that may be of equal interest, was also investigated. The available crystal structures were analyzed for this purpose only because the identification of hidden allosteric sites not apparent from the available crystal structures alone would require a large amount of conformational sampling. Such an analysis is, therefore, beyond the scope of this manuscript. The evolutionary conservation of each enzyme and binding site was calculated using ConSurf [69], and the results of each binding pocket were combined with the druggability results calculated using SiteMap. Subsequently, the physicochemical properties calculated by SiteMap were analyzed and compared with the physicochemical properties of available small-molecule inhibitors. To our knowledge, this is the first study of the shikimate pathway that considers evolutionary conservation, the identification of binding pockets, and an evaluation of druggability. We believe that this systematic study provides structural and molecular insights into the druggability and evolutionary conservation of the enzymes of this pathway and predicts whether they are suitable targets for the development of new broad-spectrum antibacterial drugs. This would open new opportunities for the development of novel inhibitors against these enzymes.

2. Results and Discussion

2.1. Study Design

In the present study, evolutionary conservation and druggability were investigated for seven enzymes of the shikimate metabolic pathway. For each protein of the pathway, we also analyzed its relevant functional role and the existence of different isozymes in the literature. The results of our analysis can be summarized in several steps that are illustrated in the workflow shown in Figure 1.

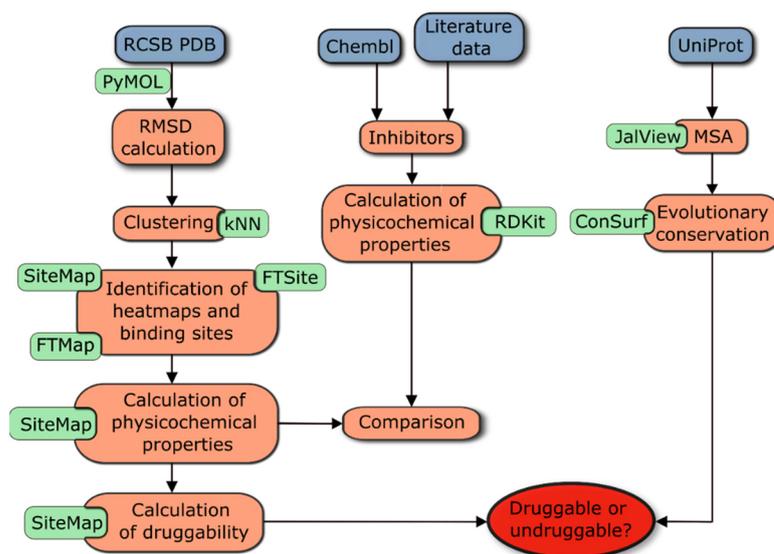


Figure 1. Summary of the workflow to determine the druggability, evolutionary conservation, and physicochemical properties of binding pockets of enzymes in the shikimate pathway.

To assess evolutionary conservation, the first step was to extract all sequences from the UniProt database and perform a multiple sequence alignment (MSA). This information was then used in the next steps to interpret the identified binding sites and druggability results. In parallel, the FTMap and FTSite servers were applied to selected enzyme structures to identify heatmaps and binding sites (including allosteric sites). This analysis

was complemented by the use of SiteMap, which identified binding pockets, calculated key physicochemical properties, and estimated druggability parameters for each binding pocket, such as DScore and SiteScore [41]. The analysis was completed by comparing these properties between representative enzymes of each class and by comparing the physicochemical properties of currently available nanomolar inhibitors of each enzyme. Complete data for all enzymes analyzed, including their physicochemical properties and druggability data for each identified site, can be found in Table S30 of the Supplementary Materials. A detailed description of the workflow procedure can be found in Materials and Methods, Section 3.1.

Four enzymes in this pathway, namely DHQS, SDH, EPSPS, and CS, each exist in only one enzyme form. In contrast, the other three enzymes, namely DAHPS, DHQase, and SK, each exist in at least two isozymes. Given that they differ in their physicochemical properties, amino acid sequences, quaternary structures, molecular mass, and occurrence in different bacteria, they were each analyzed differently. Both sequences of DAHPS isozymes were used in the MSA because they share a common 3D structure and were present in 14% of genomes analyzed [70–75]. In contrast, the two DHQase families were analyzed separately (DHQase I and II) because there is neither a unique amino acid sequence nor 3D similarity between the two [50,76]. Only the dominant form of SK, namely SK II, was used in the MSA [77]. A more detailed explanation of the properties of each isozyme, with additional details on their exact lengths, can be found in the Supplementary Materials (Table S2, Figures S4 and S5).

2.2. The Identification of Binding Sites

The binding potential of all seven enzymes in the pathway was predicted using FTMap, FTSite, and SiteMap for a set of 42 proteins from the RCS PDB. In general, for each enzyme, the number of heatmaps identified with FTMap was higher than the number of ligand-binding sites identified with SiteMap, as shown in Figure 2. This is a direct result of differences in the methodology used by both programs to calculate these sites, as well as differences in the basic definitions of heatmaps and ligand-binding sites. Heatmaps, also known as hot spots, are relatively small regions of the binding surface that contribute to a large fraction of the binding energy [78,79]. Fesik and colleagues [35] found that small organic compounds bind almost exclusively to well-defined, localized regions of proteins, regardless of their affinity. The identification of hotspots is therefore important for the study of macromolecule–ligand interactions, the identification of binding sites, the determination of druggability, and the determination of the input for fragment-based ligand discovery (FBLD) [39]. On the other hand, ligand-binding sites are regions that are larger and may contain multiple hotspots [80]. A tight-binding inhibitor may therefore bind to multiple hotspots simultaneously. It is not guaranteed that a high hit rate means that it is easy to identify a small-molecule inhibitor with sufficient binding affinity. Many proteins have binding hotspots that bind fragments with high affinity but are considered difficult drug targets [35,79].

On average, FTMap and SiteMap identified 6.1 ± 1.9 and 10.5 ± 1.5 binding sites and heatmaps, respectively, that were located at the substrate- (orthosteric sites), the cofactor-binding site, or at other parts of the surface (potential allosteric sites) (Figure 2). In contrast to SiteMap and FTMap, where multiple spots were detected, FTSite identified only three binding sites per enzyme. The number of potential heatmaps and binding sites is highly dependent on the shape and physicochemical properties of the cavities on the protein surface. These factors are also incorporated into the equation used by SiteMap to calculate the two druggability scores, SiteScore and DScore (see the experimental section) [39,41]. For most enzymes, the number of binding sites identified by SiteMap depends on the molecular weight of the enzymes and also correlates positively with the number of heatmaps identified by FTMap. An exception is CS, for which SiteMap identified additional binding sites in the C-terminal part of the enzyme that were not identified by FTSite.

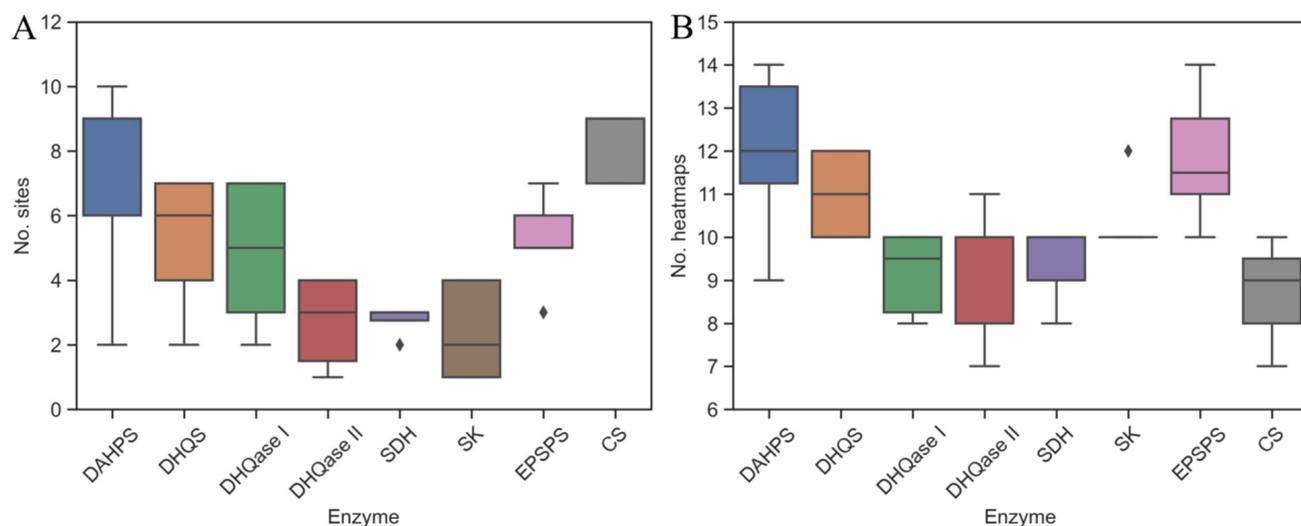


Figure 2. Identification of binding pockets: (A) SiteMap and (B) FTMap each identified several binding sites and clusters per enzyme.

2.3. Evolutionary Conservation

Figure 3 shows the projections of conservation scores obtained after MSA of sequences from different pathogenic bacteria. The scores were projected onto the protein structure of each of the seven enzymes and classified by the ConSurf server into classes ranging from 9 (highest conservation) to 1 (highest variability). The most variable positions (grade 1) are colored turquoise, the moderately conserved positions (grade 5) are colored white, and the most conserved positions (grade 9) are colored maroon. We must clarify that a conservation score calculated by ConSurf is a relative measure of evolutionary conservation at each sequence position and does not necessarily mean that the highest score indicates 100% conservation. These figures provide an overview of the structure and surface area of these enzymes and also illustrate the regions with highly conserved residues. In general, most of the conserved residues are located inside the enzyme at the substrate-binding site. This is no surprise, as the degree of evolutionary conservation is often indicative of the importance of the position in maintaining the structure and/or function of the protein. Moreover, it has already been reported that selective pressure acting on the substrate-binding site generally spreads to the rest of the protein via residue–residue contacts [81]. Thus, a gradient of conservation can be observed, such that the closer a residue is to the binding site, the more conserved it is, which is particularly observed in surface representations of enzymes (B and C). Small areas of conserved residues can be observed in the other parts of each protein, but they are small and therefore not important from a drug design perspective. An exception is CS, where a larger conserved region is found at the C-terminal end of the enzyme.

Next, we analyzed the percentage of conserved and highly conserved residues (scores 8 and 9) in each enzyme, structural motif, and binding pocket. The results are shown in Figure 4. The first histogram (A) shows that most of the conserved residues are located on the α -helices, which are inside the enzyme. Exceptions are DAHPS, where most conserved residues are found on the loops, and CS, where these residues are evenly distributed between helices and loops.

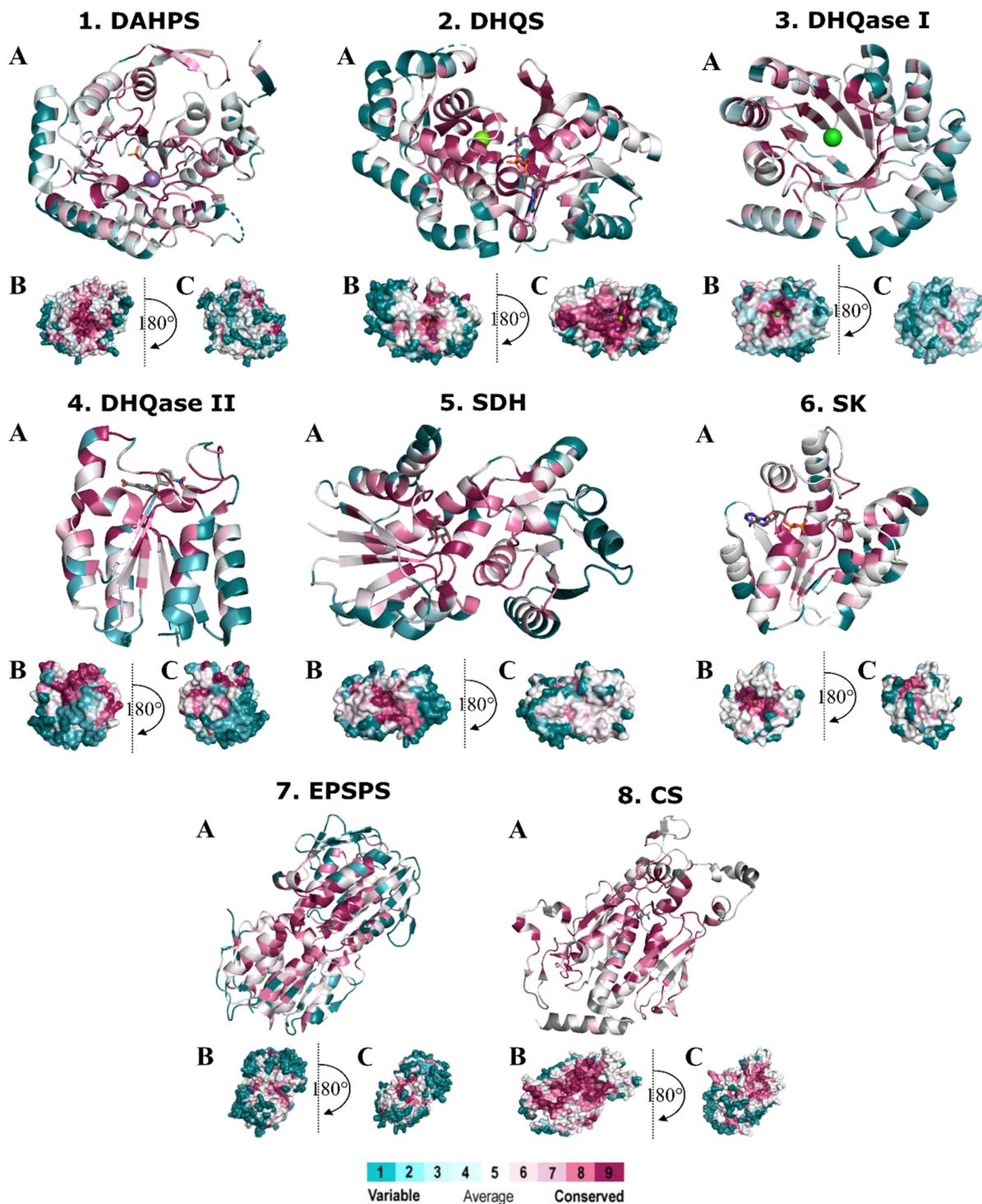


Figure 3. Spacefill (A) and surface (B,C) representation of the representative X-ray monomer structures of the individual enzymes in the shikimate pathway with the degrees of conservation mapped onto the monomer structure. Amino acids are colored according to their degree of conservation using the color-coding bar, with turquoise-through-maroon indicating variable-through-conserved residues. Enzymes: 1. DAHPS from *E. coli* (PDB: 1GG1, ligand: 2-phosphoglycolic acid), 2. DHQS from *A. baumannii* (PDB: 5EKS, ligand: NAD), 3. DHQase I from *S. enterica* (PDB: 4CNN), 4. DHQase II from *M. tuberculosis* (PDB: 4KIW, ligand: 5-[(3-nitrobenzyl)amino]benzene-1,3-dicarboxylic acid), 5. SDH from *M. tuberculosis* (PDB: 4P4G, ligand: shikimic acid), 6. SK II from *M. tuberculosis* (PDB: 2IYQ, ligand: shikimic acid, NAD), 7. EPSP synthase from *E. coli* (PDB: 2AA9, ligand: shikimic acid), 8. CS synthase from *M. tuberculosis* (PDB: 2QHF, ligand: nicotinamide).

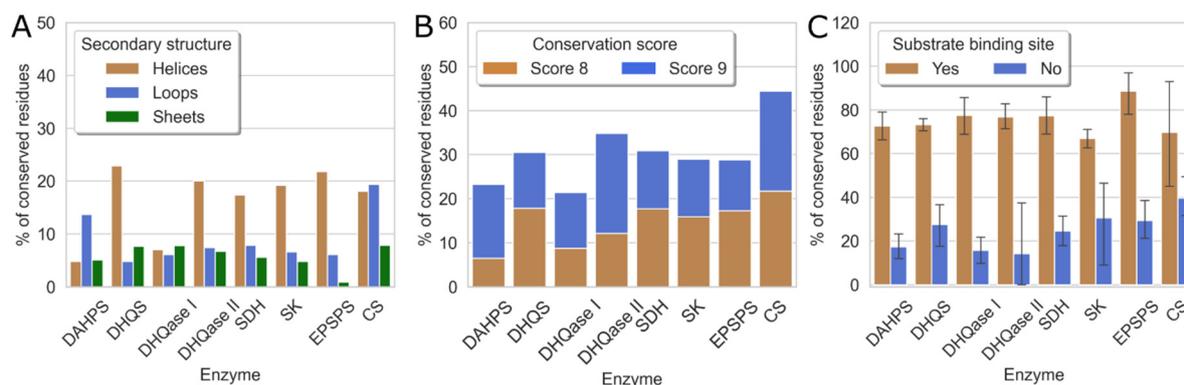


Figure 4. The proportion of evolutionarily conserved residues for individual enzymes. (A) Distribution of grade 8 and 9 residues among α -helices, β -sheets, and loops; (B) distribution of residues with conservation scores 8 and 9; and (C) binding sites identified by SiteMap.

A second, stacked histogram (B) shows that all enzymes in the shikimate pathway are evolutionarily well conserved, with conservation scores of 8 and 9 ranging from 20% to 45%, consistent with their functional importance to the bacteria. The evolutionarily least conserved are DAHPS and DHQase I, which have 23% and 21% conserved residues, respectively. The lower conservation of DAHPS is not surprising because the sequences taken for MSA were distributed between both enzyme classes and have only 10% sequence identity [70–72]. On the other hand, CS is the most evolutionarily conserved of all enzymes in the shikimate pathway and contains 44% of residues with high conservation scores.

This is also partially evident from the third histogram (C), which shows the evolutionary conservation of binding sites. These results indicate that substrate- and cofactor-binding sites have significantly higher conservation scores compared to other binding sites. It is also clear that substrate-binding sites in the pathway have a similar proportion of residues, with the highest conservation scores of 8 and 9 per binding site, ranging from 60% to 90%. In contrast, for the binding sites that do not bind substrates, this proportion is only 20 to 40%. The observation that substrate-binding sites are more conserved compared with the rest of the protein is consistent with previous literature reports [82,83]. These results also suggest a possible explanation for the high evolutionary conservation of CS. Based on the last histogram, the substrate-binding site of CS is not significantly different from other enzymes in this pathway in terms of the percentage of highly conserved residues. Its overall higher evolutionary conservation is a consequence of the fact that other parts of the enzyme that do not bind the native ligand, particularly at the C-terminal end of the enzyme, are somewhat more conserved in CS than in other enzymes of the pathway, resulting in the higher evolutionary conservation of the entire CS enzyme.

Because three enzymes in this metabolic pathway, namely DHQS, SDH, and SK, require cofactors for their normal function, we further analyzed whether there is a difference in evolutionary conservation between cofactor- and substrate-binding sites. Unfortunately, we could not use the pre-existing residues from the SiteMap results because, in most cases, the cofactor- and substrate-binding sites collided and formed one large binding site. Therefore, residues within 4 Å of the cofactor or substrate were selected for the analysis. A list of the residues of the selected representatives for each enzyme can be found in the Supplementary Materials (Tables S3–S29), along with the ConSurf score and the maximum identity of each amino acid. It was found that although both the cofactor- and substrate-binding sites have high conservation scores and are therefore viable targets for antibacterial drugs, there is a significant difference in evolutionary conservation between the two sites. The cofactor-binding sites are, on average, 20–30% less conserved than the substrate-binding sites, which all contain 100% residues with scores of 8 and 9. These results, along with the potential selectivity issues arising from the fact that these cofactors are used by several other enzymes in the human body, suggest that the substrate-binding sites of DHQS, SDH, and SK are the preferred primary targets from a drug design perspective.

The only enzyme in the entire pathway reported to be controlled by allosteric feedback inhibition is DAHPS. However, in most organisms, this enzyme exists in three isozymes: aroF, aroG, and aroH, each regulated by different aromatic amino acids, namely Tyr, Phe, and Trp [84–88]. Because all three isozymes may be present simultaneously in the bacterial cell, a potential antibacterial drug would need to act simultaneously on all three isozymes with similar binding affinities. Our results show that while the orthosteric site of this enzyme is highly conserved, the allosteric site is highly variable. For example, in *E. coli*, the only conserved amino acid was S211, with a conservation value of 8, and all other residues were significantly less conserved (Supplementary Materials, Table S4). Although it is known that allosteric sites are evolutionarily less conserved because they have not been subject to direct evolutionary pressure to conserve important functional residues [89], this can be an advantage in some cases because it leads to improved selectivity and is desirable when only one species is targeted [90,91]. For example, allosteric sites of functional proteins of *M. tuberculosis* (Mtb) have been used in the study of the Mtb enzymes ornithine acetyltransferase [92] and tryptophan synthase [93]. Thus, this lower evolutionary conservation of all sites is an advantage when only one species is targeted. On the other hand, targeting allosteric sites can also lead to species differences that may hinder the broad-spectrum efficacy of a drug in different pathogenic bacterial species [90,91].

2.4. Druggability Assessment

Both ligandability and druggability mappings were examined in detail for each of the 42 proteins from the RCS PDB database using all three programs. Both FTSite and FTMap ranked the consensus clusters by the number of unbound contacts between the protein and all probes in the cluster. Representative structures of each enzyme were then superimposed, and the three highest-scoring clusters from the FTMap and FTSite analyses were selected and are shown in Figures S6 and S7 of the Supplementary Materials. Although both FTMap and FTSite are of great help in prioritizing binding sites and are also widely used for druggability assessment [33,36,94–96], we decided to use SiteMap for this purpose because it has demonstrated its effectiveness in several publications and has also been extensively validated on a set of 538 crystal structures [33,66,97–107]. The results of FTMap and FTSite were therefore complemented by the SiteMap analysis of the same PDB structures.

In general, the highest-ranked ligandable regions selected by all three programs were the orthosteric and the cofactor regions. The exception was DAHPS, where the heatmaps were distributed around the enzyme, but most of the highest-ranking clusters were in the orthosteric region regardless of the X-ray structure used. These results were also confirmed by DScore and SiteScore calculations.

Figure 5 shows that the substrate-binding sites are generally much more ligandable than the allosteric binding sites. Two black dotted lines divide the scatter plot into four regions, with DScore and SiteScore values above and below 0.8, respectively; this is a threshold proposed by SiteMap developers to identify sites that are ligandable (SiteScore) and/or druggable (DScore) [80]. Based on these results, the majority of allosteric sites have both scores below 0.8 and are located in the lower-left quadrant, indicating the least promising sites. The majority of orthosteric sites and the minority of allosteric sites have scores above 0.8, indicating their druggability. There is also a smaller number of sites in the lower right quadrant that are undruggable but have the potential to bind small molecules with high affinity.

The observed ability of all three programs to predict substrate- and cofactor-binding sites as equally favorable for noncovalent binding also suggests a possible targeting strategy in which an inhibitor targeting the substrate-binding region could readily extend to the cofactor-binding region and interact more strongly with surrounding amino acids.

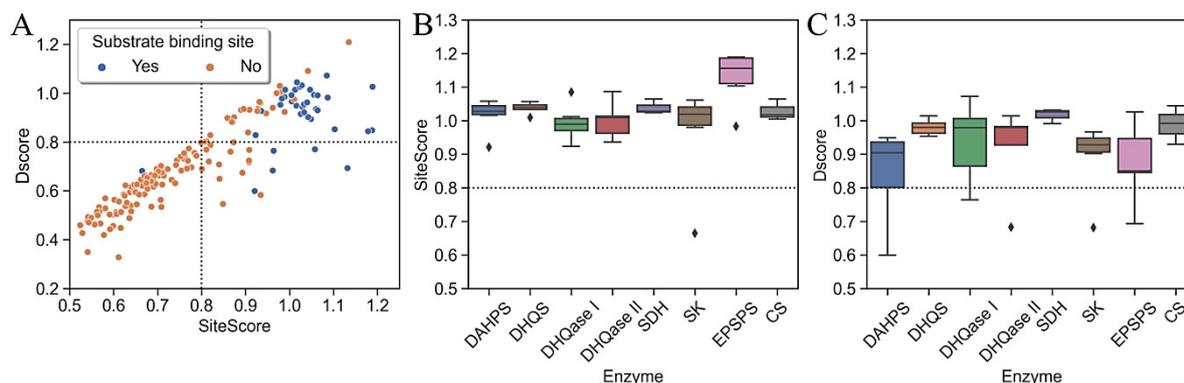


Figure 5. Distribution of ligandability and druggability parameters. (A) Scatter plot of DScore versus SiteScore numbers per type of binding site. Two black dotted lines indicate a boundary where both scores are 0.8, dividing the scatterplot into four regions, each with a different potential for druggability and ligation; (B) DScores and (C) SiteScores of substrate-binding sites per each enzyme in the pathway. Dotted lines in each boxplot indicate a reference value for the average submicromolar site.

Because one of the goals of our study was to identify binding sites other than substrate- or cofactor-binding sites, the results from FTSite and SiteMap were compared. To reduce the number of sites that could not be targeted, a threshold of 0.8 was set for SiteScore and DScore, as suggested by the SiteMap developers. This proved to be very useful because most of the undruggable binding sites were filtered out, leaving only 33% of the original sites [80]. We also found that most of the highly conserved cavities on the surface of the C-end of CS, which we mentioned earlier, were not druggable enough. Although the evolutionary conservation of protein cavities is often related to their druggability [108], most of these sites were either too small or not sufficiently enclosed. These sites were therefore filtered out and not analyzed further.

As expected, both methods largely agreed on the highest scoring sites (Supplementary Materials, Tables S5, S8, S11, S14, S17, S20, S23, S26 and S29). The binding sites that scored higher with the FTSite method largely overlapped with the binding sites that scored higher with SiteMap. However, each method also found binding sites in regions that did not bind a substrate or cofactor and that were not predicted by the other method. For example, for SDH, we identified a potential binding site in 3 of 6 crystal structures that were predicted to be druggable by SiteMap but were not identified by FTSite (Supplementary Materials, Table S17). In many cases, binding sites observed in one crystal structure and ranked highly by one of the methods were not observed in another structure of the same enzyme belonging to a different bacterial species. This observation could be a consequence of different protein conformations or differences in protein sequences leading to the lower evolutionary conservation of a single protein. Therefore, it was very difficult to decide which of these mismatched binding sites was a prospective drug binding site when only druggability results were considered.

In the context of this work, it was therefore of great importance that the druggability results were examined in conjunction with the evolutionary conservation calculations, which were very helpful in assessing whether a region of a protein was a prospective binding site or not. After using 50% of the conserved residues in a pocket as a threshold, we found that 20% of the original sites were both druggable and evolutionarily conserved, of which only two sites (SDH and CS) did not bind a substrate or cofactor. Further visual inspection revealed that the reason for the higher evolutionary conservation of these two binding regions was partial overlap with the evolutionarily well-conserved substrate-binding region and that both regions were identified as part of a larger pocket in other crystal structures of the same enzyme. The drawback of this approach, however, is that cavities that could serve as potential allosteric sites were likely filtered out because of the well-documented lower conservation of such sites [109,110].

From this, we can conclude that for all seven enzymes studied, only the substrate- or cofactor-binding regions are suitable broad-spectrum antibacterial targets and that no allosteric site can be used for this purpose.

2.5. Evaluation of Physicochemical Properties of Binding Sites

The ligandability and druggability of binding sites depend on many physicochemical parameters, such as volume, enclosure, hydrophobicity, and percentage of charged residues [80]. Drug binding pockets tend to be deeper, larger, and more complex in shape; they also tend to have high hydrophobicity, lower hydrophilicity, and high enclosure; the latter maximizes the surface area to volume ratio. The opposite is true for “hard to drug” and undruggable pockets [31,34,40,111]. Since DScore is calculated directly from the number of site points, the enclosure, and the polarity, we can further explain the differences between the druggability of orthosteric and cofactor-binding sites and the nonbinding sites by comparing their basic physicochemical properties.

These properties, which were used for evaluation, are listed in Table S30 of the Supplementary Materials. The substrate-binding sites of these enzymes are significantly larger and more enclosed compared to the allosteric sites ($p < 0.05$), which increases their DScores and SiteScores. The size of the pocket is an important descriptor because small cavities, usually consisting of <14 residues, do not promote adequate interaction between the target and the ligand [111–113]. On the other hand, substrate-binding sites were found to be more polar compared to allosteric sites ($p < 0.05$), which negatively affects their druggability. This result was not surprising since the substrates and products of each step of the pathway are also highly polar. Our results on the differences in druggability between the orthosteric and allosteric binding sites are consistent with data found in the literature [114,115].

Druggability can also be estimated from other parameters calculated indirectly from SiteMap, such as the proportion of charged residues in a binding site. It is clear from the literature that highly charged pockets are much less druggable than uncharged pockets [31,34,40,111]. Perola, therefore, proposed a threshold of 26.3% to distinguish between druggable and non-druggable pockets [111]. However, although substrate-binding sites were found to be more polar than allosteric sites, we did not detect a significant difference in the proportion of charged residues between the average orthosteric and allosteric sites ($p < 0.5$).

Most of the metrics used to evaluate the potential allosteric sites, such as DScore, the number of site points, the enclosure, and the volume, were below the thresholds that distinguish druggable from the undruggable sites defined in the literature [33,66,97–107]. We can, therefore, additionally confirm that we were unable to identify any of the allosteric pockets as candidates for drug targeting in any of the seven enzymes analyzed.

2.6. Druggability of the Substrate-Binding Sites

Next, we turned to a detailed evaluation of the orthosteric and cofactor-binding sites and an analysis of their physicochemical properties. We found that the binding sites of all seven enzymes are ligandable, as indicated by their high SiteScore values around 1; therefore, all are considered particularly promising targets for small molecules (Figure 5). The highest-scoring enzyme in terms of its mean SiteScore value was EPSPS (SiteScore = 1.13), mainly due to its high enclosure, and the lowest-scoring enzyme was DHQase I ($p < 0.05$). Although DAHPS is particularly promising for small molecule development, SiteMap classified it as difficult to drug because its DScore was not significantly different from 0.8 ($p < 0.05$). This does not mean that the site cannot bind ligands, but that it would be difficult to find high-affinity drug-like ligands for it. Other enzymes in the pathway were found to be druggable, with a DScore around the value of 1, and they are, therefore, particularly promising targets for the development of drug-like molecules.

2.7. Physicochemical Properties of Substrate-Binding Sites

To explain the lower druggability of DAHPS and to further investigate the binding sites of all other enzymes in the pathway, the physicochemical properties of each enzyme were next evaluated. Such an analysis provides medicinal chemists with important guidelines for predicting the physicochemical properties of potential drugs for the specific binding pocket. The main properties of the individual enzymes are shown in Figure 6.

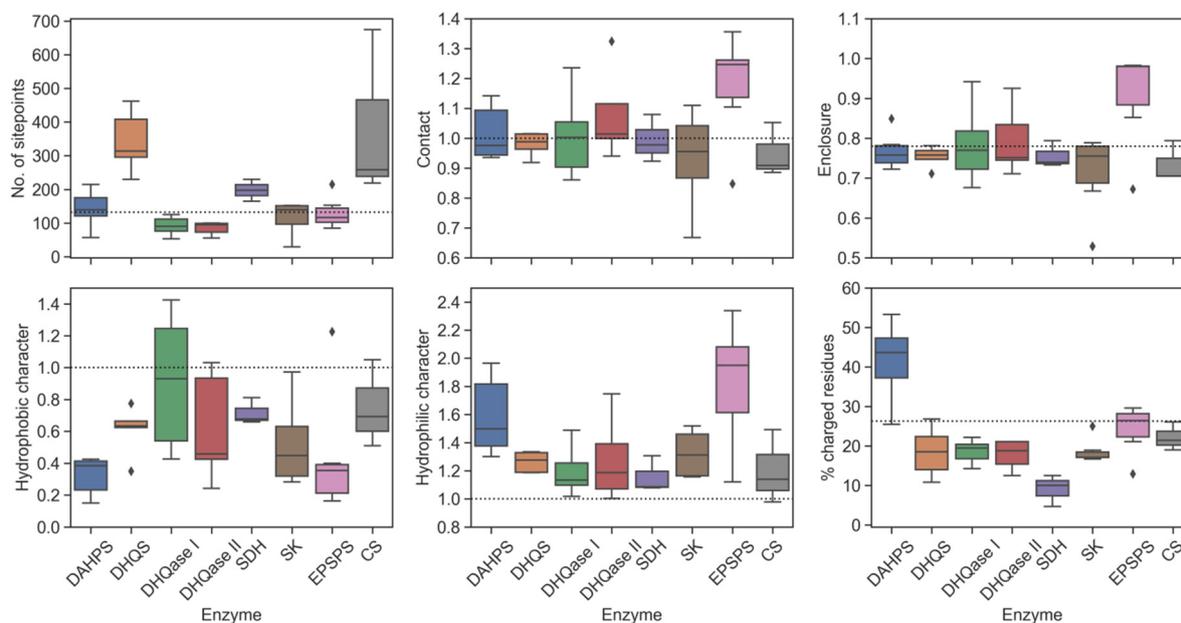


Figure 6. Key physicochemical properties of each enzyme in the shikimate pathway. The dotted line in each boxplot indicates a reference value for the average submicromolar site.

In general, the greatest differences between the enzymes were found when the volumes of the binding sites were compared. For DHQS, SDH, and CS, the volume is significantly higher than the lower limit of the average submicromolar binding sites, which is a consequence of the additional space occupied by a cofactor rather than by a substrate alone. In contrast, the binding site of SK, which uses ATP as a cofactor, was found to be similar in size to other enzymes that do not require a cofactor.

We also found that most binding sites have enclosures and strengths of van der Waals interactions with the site points (contact) comparable to the characteristics of the average tight-binding site. One enzyme that stands out when considering both properties is EPSPS, which has a very high enclosure and forms strong contacts with its binding site. These two properties increase the likelihood that an inhibitor interacts strongly with the binding site, increasing its ligandability and druggability.

Substrate- and cofactor-binding sites were found to have a higher hydrophilic character than the average druggable tight-binding site. This property is of great importance because it negatively affects the druggability of a binding pocket and also indirectly predicts the polarity of a potential ligand [30]. The higher polarity of the binding site is related to the relatively higher polarity of the substrates of these enzymes. In addition, two enzymes, DAHPS and EPSPS, tend to have an extremely hydrophilic character as a result of their highly charged substrates with two acidic groups. However, unlike EPSPS, which has a very high enclosure that increases its DScore, the enclosure of DAHPS is slightly below the average of tight-binding sites. Consequently, its DScore is at the threshold of 0.8 and it has a very high percentage of charged residues, well above the proposed threshold of 26.3%. This site should therefore be considered an undruggable drug target. Interestingly, although the binding site of EPSPS is also very polar, the percentage of charged residues is below the proposed threshold. The high SiteScore suggests that this binding site is more ligandable

compared to other sites in the pathway. On the other hand, a hypothetical submicromolar inhibitor targeting such a highly polar binding pocket should also have higher polarity, which could lead to problems in pharmacokinetic properties, such as a shorter elimination half-life [116].

From a drug-design perspective, a binding site should also have a significant number of hydrophobic regions in a pocket that can bury the drug in the pocket. Since electrostatic interactions and desolvation energies act in opposite directions, their contribution to binding potency is negligible. Therefore, hydrophobic interactions are a driving force in the interactions between a ligand and the binding site and, consequently, have the greatest influence on the strength of drug binding [80,117]. Despite their higher proportion of hydrophilic surface area compared with the average tight-binding site, we found that DHQase I, II, and CS have larger regions of hydrophobic amino acids than other enzymes, and their calculated hydrophobic character is comparable to that of the average drug binding site. For these enzymes, it should be easier to design a drug-like inhibitor that interacts with both hydrophobic and hydrophilic regions of the enzyme.

Because the physicochemical properties of each binding site predict the optimal properties of a potential tight-binding inhibitor, a comparison was made between SiteMap results and the values calculated for published inhibitors. Although the number of drug-like nanomolar inhibitors is too small to predict the druggability of each enzyme in the pathway, a comparison between the predicted properties and the existing inhibitors may provide some clues on how to modify the properties of the inhibitors to further improve their binding efficacy. Therefore, two physicochemical properties, topological polar surface area (TPSA) and hydrogen bond donor/acceptor ratio (don/acc), were calculated for the published inhibitors. These results are shown in Figure 7. Since most of the published inhibitors are derivatives or analogs of native ligands, a slight correlation between their TPSA and the hydrophilic character of the binding sites was expected and, for the most part, found. The inhibitors of highly polar binding sites of EPSPS and DAHPS, therefore, have high TPSA values, whereas the inhibitors of other enzymes have similar TPSA values. As expected, EPSPS and DAHPS inhibitors are not drug-like due to their high polarity, which is a consequence of the presence of phosphonic and carboxylic groups in these compounds. Therefore, the binding strength and the drug-likeness of the inhibitors could be further improved by changing the ratio of don/acc groups and eliminating at least one of the highly acidic groups or by preparing prodrugs. In the boxplot in Figure 7, we see that the predicted don/acc ratio of an ideal inhibitor is higher than the actual ratio for the DAHPS, DHQS, and CS enzymes. Thus, an inhibitor with a higher don/acc ratio would be a better match for these binding sites.

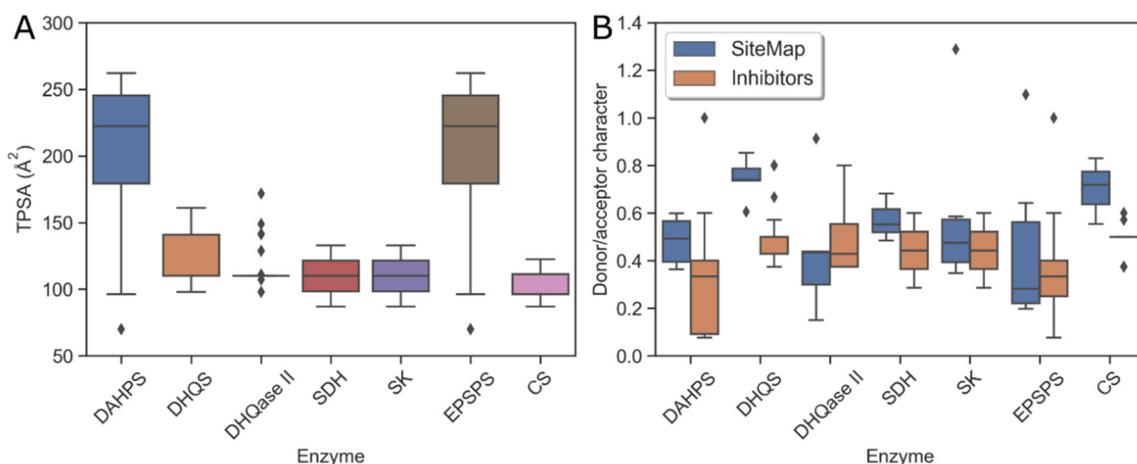


Figure 7. Physicochemical properties of the substrate-binding sites and the available inhibitors. (A) Median TPSA of inhibitors per enzyme; (B) proton donor/acceptor character of inhibitors compared to binding sites per enzyme.

3. Materials and Methods

3.1. Study Design

To assess evolutionary conservation, the first step was to extract all sequences and other data for each sequence from the UniProt database. All filtering and analyses of the data were performed using KNIME [118], and only representative sequences from pathogenic bacteria were selected. Subsequently, MSA of the selected sequences was performed, and a phylogenetic tree (PT) was constructed. MSA and PT are prerequisites for the calculation of evolutionary conservation, which was performed using the free online bioinformatics tool ConSurf server [69,119–121]. This information was then used in the next steps to interpret the identified binding sites and druggability results. Because many X-ray structures are publicly available, the 3D structures of each bacterial protein were next extracted from RCS PDB. The structures of each enzyme were then aligned, and root-mean-square deviation (RMSD) was calculated, followed by k-means clustering. Representative structures from each of the six clusters were selected based on resolution, average B-factor, and the number of missing residues in each PDB structure. Structures were then prepared using the Protein Preparation Wizard implemented in the Schrödinger Suite, which is described in more detail in Section 3.4. To identify all heatmaps and binding sites (including allosteric sites) in each protein, the FTMap and FTSite servers were applied to selected representatives. This analysis was complemented by the use of SiteMap, which uses the interaction energies between the protein and grid probes to locate energetically favorable sites, identify binding pockets, calculate key physicochemical properties, and estimate druggability parameters, such as DScore and SiteScore, of each binding pocket [41]. The analysis was completed by comparing these properties between representative enzymes of each class and by comparing the physicochemical properties of the currently available nanomolar inhibitors of each enzyme. Complete data for all enzymes analyzed, including their physicochemical properties and druggability data for each identified site, are provided in Table S30 of the Supplementary Materials.

3.2. Identification of Hotspots and Binding Sites with FTMap and FTSite

FTMap [39] is a computational mapping server that identifies binding hotspots of macromolecules using 16 small organic probe molecules of different sizes, shapes, and polarities distributed on the molecular surface. Hotspots are smaller regions of proteins that contribute significantly to the binding of a drug to the binding site, and their strength determines the druggability of a site. They are very important for fragment-based ligand discovery (FBLD), as the ligand moieties that interact with them are essential for binding. FTSite is a second server that identifies ligand binding sites by using consensus sites determined via the FTMap server to identify and rank binding sites. Similar to FTMap, it uses molecular probes to map the macromolecular surface and finds the most favorable positions for each probe type. Probes are then clustered and ranked by the number of nonbonded contacts between the protein and all probes in the consensus cluster rather than by the number of probe clusters. The following probes are used by each server: Acetamide, acetonitrile, acetone, acetaldehyde, methylamine, benzaldehyde, benzene, isobutanol, cyclohexane, *N,N*-dimethylformamide, dimethyl ether, ethanol, ethane, phenol, isopropanol, and urea [39].

All prepared PDB files were uploaded to the FTMap and FTSite servers for mapping. Upon completion, the results were downloaded as a Pymol [122] session file (.pse) for further analysis. Amino acids within 4 Å of each cluster were selected and exported in .pdb format. PDBBest [123], a freely available platform for the manipulation, filtering, and normalization of biomolecules, was used to extract residue numbers that were further processed using KNIME [124].

3.3. Druggability Assessment with SiteMap

SiteMap [41] is software used for binding site identification and evaluation; it is implemented in Schrödinger's Maestro [85]. We chose SiteMap because it has been shown to

be effective in a number of publications on various proteins [33,66,97–107]. A comprehensive validation of SiteMap was published using 538 crystal structures from the PDBbind database [80]. In this study, SiteMap accurately identified 86% of the highest-scoring binding sites. In addition, it was found that the size of the binding site, as measured by the number of site points found, and the relative openness of the site, as measured by the exposure and enclosure properties, were the most important terms for distinguishing binding sites from nonbinding sites.

SiteMap uses the interaction energies between the protein and the grid probes to find energetically favorable sites. In a three-step process, a program generates a grid of points (site points) and then uses their energetic properties to score the sites. During this process, various physicochemical properties are calculated: size, volume, degree of enclosure or exposure, degree of contact, hydrophobic or hydrophilic character, and their balance and hydrogen bonding properties (acceptors/donors), all of which are listed in Table 1 along with their reference values for average submicromolar inhibitors. Exposure and enclosure provide different estimates of how open a binding site is to the solvent. The lower the exposure value and the higher the enclosure, the better. The hydrophobic/hydrophilic balance measures the relative hydrophobic and hydrophilic character of the site. The donor/acceptor character indicates the extent to which a potential ligand can donate rather than accept hydrogen bonds [80].

Table 1. Properties calculated by SiteMap [41] along with the average property score for a submicromolar site. Data from [80].

Property	Lower Limit ¹
Number of site points	132
Site score	1 (0.8)
DScore	1 (0.8)
Exposure	0.49
Enclosure	0.78
Contact	1.0
Phobic, Philic ²	1.0
HL balance ³	1.6
Donor/acceptor	0.76

¹ Numbers are calculated for an average submicromolar site; ² Phobic, Philic = Hydrophobic property, Hydrophilic property; ³ HL balance = Hydrophilic/lipophilic balance.

In addition, SiteScore and DScore were also calculated to evaluate overall ligandability and druggability properties.

SiteScore predicts whether a binding site could be a drug binding site. The developers suggest a value of 0.80 to distinguish sites that bind ligands from those that are not known to do so. Thus, a value greater than 1 indicates a particularly promising site, as seen in Equation (1) [80].

$$\text{SiteScore} = 0.0733\sqrt{n} + 0.6688e - 0.20p \quad (1)$$

where

n = the number of site points found for the site, capped at 100;

e = the degree of enclosure score;

p = the hydrophilic score computed for the site, capped at 1.0 to limit the impact of hydrophilicity in charged and highly polar sites.

DScore is a druggability score that includes terms that promote ligand binding but offsets them with a term that penalizes increasing hydrophilicity. Again, the number of site points is limited to 100, but the score for hydrophilicity is not capped. See Equation (2).

$$\text{DScore} = 0.094\sqrt{n} + 0.60e - 0.324p \quad (2)$$

where

n = the number of site points found for the site, capped at 100;

e = the degree of enclosure of the site;

p = the hydrophilic score computed for the site.

To facilitate the comparison of the three-dimensional structures of all seven enzymes in the shikimate pathway, the prepared proteins were first aligned using the Schrödinger's Protein Structure Alignment utility in Maestro [125]. Proteins were then prepared according to the procedure described below. SiteMap was applied to all selected enzymes to identify up to 10 potential binding sites with the highest score. Sites were retained if they comprised at least 15 site points per reported site. The narrower definition of hydrophobicity was used, along with a fine grid (0.35 Å). Site maps that were 4 Å or more from the nearest site points were truncated. The analysis of SiteMap calculations was performed using KNIME [124], extracting physicochemical property data and residue numbers for further processing.

3.4. Protein Acquisition and Preparation

Selected 3D structures were imported from the Protein Data Bank into Maestro [125]. To prevent regions between protein monomers from forming unphysical sites that give good results but exist only in the crystal lattice and not in solution, only one monomer chain was retained, usually chain A. Structures were then prepared using the Protein Preparation Wizard implemented in the Schrödinger Suite [125]. In each PDB structure, ligands, waters, and other co-crystallized molecules were removed except for the Mg cofactor, if present. Bond orders were automatically assigned, hydrogens were added, selenomethionines were converted to methionines, missing side chains were added, ligands were removed, disulfide bridges were created when possible, waters beyond 5 Å radius of heteroatoms were added, and heteroatoms were protonated at pH 7.0. The impref utility was used to perform constrained minimization of the protein with a maximum root-mean-square deviation (RMDS) of 0.30 Å.

3.5. Superposition of 3D Structures and Selection of Representative Crystal Structures

All 3D structures available for each enzyme were taken from the Protein Data Bank (PDB) and prepared using the procedures described in Section 2.3. Water was removed, and only one chain, mostly chain A, was retained for analysis. All structures were then aligned, and the RMSD was calculated for each pair. The values were exported to KNIME in the form of a 2D matrix for further processing. Only bacterial enzymes without induced residue mutations were selected, and then the k-means algorithm was used to generate six representative clusters. The selection of the representative structure from each cluster was then based on resolution and the extent to which some protein chains were incomplete.

3.6. Multiple Sequence Alignment

The protein sequences of the individual enzymes of all species were obtained from UniProt [126]. Data on the total number of sequences extracted per enzyme and the distribution among kingdoms are given in Table S1 of the Supplementary Materials. Outliers that had only 60% residues compared with the average length of the protein family were filtered out, and then a unique representative was selected for each human pathogenic bacterium. A list of the number of human pathogenic microorganisms was obtained from the Kyoto Encyclopedia of Genes and Genomes [127].

The protein sequences of the human pathogens were then subjected to multiple sequence alignment with MuscleWS in Jalview software [128,129] using the default settings. After converting the alignment to Fasta file format, a phylogenetic tree was calculated based on the neighbor-joining method using similarity scores calculated by the blocks substitution matrix 62 (BLOSUM 62), which measures the evolutionary relationship between each pair of sequences in the alignment [130]. After converting the MSA and phylogenetic tree to the Clustal Alignment (.aln) and Newick (.txt) format files, respectively, the multiple alignment file and phylogenetic tree were subsequently used by the ConSurf server [69,119–121].

3.7. Calculation of Evolutionary Conservation

The ConSurf server [69,119–121] is a bioinformatics tool for estimating the evolutionary conservation of amino/nucleic acid positions in a protein/DNA/RNA molecule based on the phylogenetic relationships between homologous sequences. The program uses either an empirical Bayesian method or a maximum likelihood method [131,132] to estimate the evolutionary rate of the protein sequence, which is then projected onto the 3D structure of a selected enzyme.

Both of the .aln files from the MSA and the phylogenetic tree were used by the ConSurf server to calculate the conservation scores, which were then projected onto the X-ray structure of the selected enzyme. The evolutionary conservation of the selected enzymes in pathogenic bacteria was then calculated using the empirical Bayesian calculation method [131,132] with a default substitution model. The conservation scores calculated by ConSurf are a relative measure of evolutionary conservation at each sequence site in the target chain. The obtained continuous conservation scores were mapped onto the X-ray structure of a protein and divided into a discrete scale of nine grades for visualization, from the most-variable positions (grade 1), which are colored turquoise, through intermediately conserved positions (grade 5), which are colored white, to the most-conserved positions (grade 9), which are colored maroon. Results were downloaded from the ConSurf website as a .pdb file containing data for 3D coordinates and conservation scores. Figures mapping the conservation scores to the protein structures were subsequently generated using Pymol [122] for one representative of each enzyme and are shown in Figure 6. A list of ConSurf grades was also obtained and imported into KNIME to extract the names of all residues, along with their conservation scores and identity thresholds.

3.8. Calculation of TPSA

Information on inhibitors for each enzyme was obtained from the literature and the ChEMBL database [133]. Inhibitors were first prepared using LigPrep, which is implemented in the Schrödinger Suite [125]. Conformations of the molecules were generated using the OPLS4 force field and ionized using Epik [134,135] at a target pH of 7 ± 2 . The molecules were desalted, and the tautomers were generated. The TPSA was then calculated using QikProp, which is also implemented in the Schrödinger Suite [125].

3.9. Statistical Analysis

To compare the means of DScore and SiteScore between samples, the Kolmogorov–Smirnov test was used to test whether the data were normally distributed, followed by the Mann–Whitney U test. The difference between the two groups was statistically significant when the *p*-value was less than 0.05.

3.10. Generation of Figures

The graphical representation of the three-dimensional structures of proteins was performed using Pymol [123]. Boxplots and stackplots were generated using the Seaborn [136] and Matplotlib [137] libraries in Jupyter notebook [138].

4. Conclusions

The exploitation of important but underexplored antibacterial targets is one of the ways to address the urgent need for new antibacterial agents. In the absence of clinically relevant antibacterial drugs against any of the enzymes of the shikimate pathway, we hypothesized that these targets are simply not ligandable or druggable enough to produce a drug-like inhibitor against any of them, given the 70 years of research in this field.

In the present study, we report an *in silico* assessment of the evolutionary conservation, ligandability, and, most importantly, druggability of the enzymes belonging to the shikimate pathway. To our knowledge, this is the first and most comprehensive study using a combination of all three features for all of these enzymes.

Because mutations and resulting drug resistances play an important role in antibacterial drug development, a validation of antibacterial targets should involve a combination of estimations of druggability, evolutionary conservation, and physicochemical properties of individual binding sites. Using this approach, we mapped the conserved residues in the 3D structures of individual enzymes using ConSurf and applied these results to the assessment of discovered binding pockets; this was performed using FTMap, FTSite, and SiteMap. We have shown that only substrate-binding sites in various pathogenic bacteria are both evolutionarily conserved and ligandable. With the exception of DAHPS, these sites can be targeted by drug-like molecules and are all attractive targets for pharmacological modulation. All of these sites have higher polarity than the average tight-binding site, with DAHPS and EPSPS having particularly high polarity. Because the physicochemical properties of a binding site influence the properties of inhibitors targeting the site, we predict that any drug targeting EPSPS would also be highly polar, which was also observed in the analysis of existing inhibitors targeting EPSPS.

Although the focus of our workflow presented here was exclusively on the development of broad-spectrum antibacterial agents, the same approach could easily be applied to the analysis of the same enzymes as narrow-spectrum antibacterial targets. For example, 42% of all identified sites that are not orthosteric were found to be druggable. Because the evolutionary conservation of protein cavities is often related to their druggability, at least some of these cavities may be conserved in one bacterial species and not in the others.

We believe that this systematic study provides a sufficient structure-based rationale to accelerate the exploration of these underexplored targets. In addition, our workflow presented here may open new opportunities for the development of new antibacterial drugs targeting antibacterial agents that are not strictly related to the shikimate pathway.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/antibiotics11050675/s1>, Table S1: Sequences that were extracted from UniProt; Table S2: Average length and mass of DAHPS isoforms together with the number of sequences from different pathogenic bacteria found in the UniProt database; Table S3: Highly conserved residues (conservation score 9), conserved residues (conservation score 8), and variable residues (conservation scores 1–3) of DAHPS from *E. coli* (PDB: 1GG1), identified by ConSurf server; Table S4: Amino acids of PEP, E4P, and allosteric binding sites of DAHPS from *E. coli* (PDB: 1GG1), with their ConSurf scores and maximum identity; Table S5: Amino acid residues and the corresponding clusters of DAHPS from *E. coli* (PDB: 1GG1) identified by FTMap and SiteMap; Table S6: Highly conserved residues (conservation score 9), conserved residues (conservation score 8), and variable residues (conservation scores 1–3) of DHQS from *A. baumannii* (PDB: 5EKS) identified by ConSurf server; Table S7: Amino acids of DHQ and NAD binding sites of DHQS from *A. baumannii* (PDB: 5EKS), with their ConSurf scores and maximum identity; Table S8: Amino acid residues and the corresponding clusters of the DHQS from *A. baumannii* (PDB: 5EKS) identified by FTMap and SiteMap; Table S9: Highly conserved residues (conservation score 9), conserved residues (conservation score 8), and variable residues (conservation scores 1–3) of the DHQase I from *S. enterica* (PDB: 4CNN) identified by the ConSurf server; Table S10: Amino acids of the DHQase I binding sites from *S. enterica* (PDB: 4CNN) identified by the ConSurf server, with their ConSurf scores and maximum identity; Table S11: Amino acid residues and the corresponding clusters of DHQase I from *S. enterica* (PDB: 4CNN) identified by FTMap and SiteMap; Table S12: Highly conserved residues (conservation score 9), conserved residues (conservation score 8), and variable residues (conservation scores 1–3) of the DHQase II from *M. tuberculosis* (PDB: 4KIW), identified by the ConSurf server; Table S13: Amino acids of 3-dehydroquinate binding site of DHQase II from *M. tuberculosis* (PDB: 4KIW), with their ConSurf scores and maximum identity; Table S14: Amino acid residues and the corresponding clusters of DHQase II from *M. tuberculosis* (PDB: 4KIW), identified by FTMap and SiteMap; Table S15: Highly conserved residues (conservation score 9), conserved residues (conservation score 8), and variable residues (conservation scores 1–3) of the SDH from *M. tuberculosis* (PDB: 4P4G), identified by the ConSurf server; Table S16: Amino acids of the shikimate and NADPH binding sites of SDH from *M. tuberculosis* (PDB: 4P4G), with their ConSurf scores and maximum identity; Table S17: Amino acid residues and the corresponding clusters of SDH from *M. tuberculosis* (PDB: 4P4G), identified by FTMap and SiteMap; Table S18: Highly conserved residues (conservation

score 9), conserved residues (conservation score 8), and variable residues (conservation scores 1–3) of SK II from *M. tuberculosis* (PDB: 2IYQ), identified by ConSurf server; Table S19: Amino acids of the shikimate and ATP binding site of the SK II from *M. tuberculosis* (PDB: 2IYQ) with their ConSurf scores and maximum identity; Table S20: Amino acid residues and the corresponding clusters of SK from *M. tuberculosis* (PDB: 2IYQ) identified by FTMap and SiteMap; Table S21: Highly conserved residues (conservation score 9), conserved residues (conservation score 8), and variable residues (conservation scores 1–3) of SK II from *M. tuberculosis* (PDB: 2IYQ), identified by ConSurf server; Table S22: Amino acids of the shikimate and ATP binding site of the SK II from *M. tuberculosis* (PDB: 2IYQ) with their ConSurf scores and maximum identity; Table S23: Amino acid residues and the corresponding clusters of SK from *M. tuberculosis* (PDB: 2IYQ) identified by FTMap and SiteMap; Table S24: Highly conserved residues (conservation score 9), conserved residues (conservation score 8), and variable residues (conservation scores 1–3) of the EPSPS from *E. coli* (PDB: 2AA9) identified by the ConSurf server; Table S25: Amino acids of shikimate and PEP binding sites of EPSPS from *E. coli* (PDB: 2AA9) with their ConSurf scores and maximum identity; Table S26: Amino acid residues and the corresponding clusters of EPSPS from *E. coli* (PDB: 2AA9) identified by FTMap and SiteMap; Table S27: Highly conserved residues (conservation score 9), conserved residues (conservation score 8), and variable residues (conservation scores 1–3) of CS from *M. tuberculosis* (PDB: 2QHF) identified by the ConSurf server; Table S28: Amino acids of EPSP binding site of CS from *M. tuberculosis* (PDB: 2QHF) active and binding sites with their ConSurf scores and maximum identity; Table S29: Amino acid residues and the corresponding clusters of the CS from *M. tuberculosis* (PDB: 2QHF) identified by FTMap and SiteMap; Table S30: Druggabilities and important physicochemical characteristics of the detected binding sites; Figure S1: The number of inhibitors of the shikimate pathway enzymes; Figure S2: Inhibitors 1–10 of bacterial enzymes belonging to the shikimate pathway; Figure S3: Biosynthetic steps of the shikimate metabolic pathway; Figure S4: Comparison of protein lengths between type I and II DHQase counted in the number of amino acids that constitute each protein; Figure S5: Comparison of protein lengths between type I and II SK counted in the number of amino acids that constitute each protein; Figure S6: Surface representation of heatmaps and binding sites from 1. DAHPS, 2. DHQS, 3. DHQase I, and 4. DHQase II, identified by (a) FTMap, (b) FTSite, and (c) SiteMap, mapped onto each X-ray structure; Figure S7: Surface representation of heatmaps and binding sites of 1. SDH, 2. SK II, 3. EPSP, and 4. CS identified by (a) FTMap, (b) FTSite, and (c) SiteMap, mapped onto each X-ray structure; Figure S8: Surface representation of heatmaps and binding sites of 1. SDH, 2. SK II, 3. EPSP, and 4. CS identified by (a) FTMap, (b) FTSite, and (c) SiteMap, mapped onto X-ray structure of each; Figure S9: Differences in important physicochemical properties between the substrate and the allosteric binding sites. References [139–149] are cited in the Supplementary Materials.

Funding: This research was funded by the Slovenian Research Agency—ARRS (core financing P1-0208).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Uchil, R.R.; Kohli, G.S.; Katekhaye, V.M.; Swami, O.C. Strategies to combat antimicrobial resistance. *J. Clin. Diagn. Res. JCDR* **2014**, *8*, ME01–ME04. [[CrossRef](#)] [[PubMed](#)]
2. Cattoir, V.; Felden, B. Future antibacterial strategies: From basic concepts to clinical challenges. *J. Infect. Dis.* **2019**, *220*, 350–360. [[CrossRef](#)] [[PubMed](#)]
3. Pati, N.B.; Vishwakarma, V.; Selvaraj, S.K.; Dash, S.; Saha, B.; Singh, N.; Suar, M. *Salmonella typhimurium* TTSS-2 deficient Mig-14 mutant shows attenuation in immunocompromised mice and offers protection against wild-type *Salmonella typhimurium* infection. *BMC Microbiol.* **2013**, *13*, 236. [[CrossRef](#)] [[PubMed](#)]
4. Stritzker, J.; Janda, J.; Schoen, C.; Taupp, M.; Pilgrim, S.; Gentschev, I.; Schreier, P.; Geginat, G.; Goebel, W. Growth, virulence, and immunogenicity of *Listeria monocytogenes* aro mutants. *Infect. Immun.* **2004**, *72*, 5622–5629. [[CrossRef](#)] [[PubMed](#)]

5. Nishioka, Y.; Demerec, M.; Eisenstark, A. Genetic analysis of aromatic mutants of *Salmonella typhimurium*. *Genetics* **1967**, *56*, 341–351. [[CrossRef](#)] [[PubMed](#)]
6. Foulongne, V.; Walravens, K.; Bourg, G.; Boschioli, M.L.; Godfroid, J.; Ramuz, M.; O’Callaghan, D. Aromatic compound-dependent *Brucella suis* is attenuated in both cultured cells and mouse models. *Infect. Immun.* **2001**, *69*, 547–550. [[CrossRef](#)]
7. McArthur, J.D.; West, N.P.; Cole, J.N.; Jungnitz, H.; Guzman, C.A.; Chin, J.; Lehrbach, P.R.; Djordjevic, S.P.; Walker, M.J. An aromatic amino acid auxotrophic mutant of *Bordetella bronchiseptica* is attenuated and immunogenic in a mouse model of infection. *Fems Microbiol. Lett.* **2003**, *221*, 7–16. [[CrossRef](#)]
8. Fittipaldi, N.; Harel, J.; D’Amours, B.; Lacouture, S.; Kobisch, M.; Gottschalk, M. Potential use of an unencapsulated and aromatic amino acid-auxotrophic *Streptococcus suis* mutant as a live attenuated vaccine in swine. *Vaccine* **2007**, *25*, 3524–3535. [[CrossRef](#)]
9. Smith, D.A.; Parish, T.; Stoker, N.G.; Bancroft, G.J. Characterization of auxotrophic mutants of *Mycobacterium tuberculosis* and their potential as vaccine candidates. *Infect. Immun.* **2001**, *69*, 1142–1150. [[CrossRef](#)]
10. Bentley, R. The shikimate pathway—A metabolic tree with many branches. *Crit. Rev. Biochem. Mol. Biol.* **1990**, *25*, 307–384. [[CrossRef](#)]
11. Gibson, F.; Pittard, J. Pathways of biosynthesis of aromatic amino acids and vitamins and their control in microorganisms. *Bacteriol. Rev.* **1968**, *32*, 465–492. [[CrossRef](#)]
12. Davis, B. Aromatic biosynthesis 1. The role of shikimic acid. *J. Biol. Chem.* **1951**, *191*, 315–325. [[CrossRef](#)]
13. Roberts, F.; Roberts, C.W.; Johnson, J.J.; Kyle, D.E.; Krell, T.; Coggins, J.R.; Coombs, G.H.; Milhous, W.K.; Tzipori, S.; Ferguson, D.J.P.; et al. Evidence for the shikimate pathway in apicomplexan parasites (Vol 393, Pg 801, 1998). *Nature* **1998**, *395*, 306. [[CrossRef](#)]
14. Herrmann, K.M.; Weaver, L.M. The shikimate pathway. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1999**, *50*, 473–503. [[CrossRef](#)] [[PubMed](#)]
15. Coracini, J.D.; de Azevedo, W.F. Shikimate kinase, a protein target for drug design. *Curr. Med. Chem.* **2014**, *21*, 592–604. [[CrossRef](#)] [[PubMed](#)]
16. Steinrucken, H.; Amrhein, N. The Herbicide glyphosate is a potent inhibitor of 5-enolpyruvyl-shikimic-acid 3-phosphate synthase. *Biochem. Biophys. Res. Commun.* **1980**, *94*, 1207–1212. [[CrossRef](#)]
17. Sikorski, J.A.; Gruys, K.J. Understanding glyphosate’s molecular mode of action with EPSP synthase: Evidence favoring an allosteric inhibitor model. *Acc. Chem. Res.* **1997**, *30*, 2–8. [[CrossRef](#)]
18. Steinrucken, H.; Amrhein, N. 5-enolpyruvylshikimate-3-phosphate synthase of klebsiella-pneumoniae. 2. Inhibition by glyphosate [N-(phosphonomethyl)glycine]. *Eur. J. Biochem.* **1984**, *143*, 351–357. [[CrossRef](#)]
19. Roberts, C.W.; Roberts, F.; Lyons, R.E.; Kirisits, M.J.; Mui, E.J.; Finnerty, J.; Johnson, J.J.; Ferguson, D.J.P.; Coggins, J.R.; Krell, T.; et al. The shikimate pathway and its branches in apicomplexan parasites. *J. Infect. Dis.* **2002**, *185*, S25–S36. [[CrossRef](#)]
20. Mir, R.; Jallu, S.; Singh, T.P. The shikimate pathway: Review of amino acid sequence, function and three-dimensional structures of the enzymes. *Crit. Rev. Microbiol.* **2015**, *41*, 172–189. [[CrossRef](#)]
21. Ducati, R.G.; Basso, L.A.; Santos, D.S. Mycobacterial shikimate pathway enzymes as targets for drug design. *Curr. Drug Targets* **2007**, *8*, 423–435. [[CrossRef](#)] [[PubMed](#)]
22. Nunes, J.E.S.; Duque, M.A.; de Freitas, T.F.; Galina, L.; Timmers, L.F.S.M.; Bizarro, C.V.; Machado, P.; Basso, L.A.; Ducati, R.G. *Mycobacterium tuberculosis* shikimate pathway enzymes as targets for the rational design of anti-tuberculosis drugs. *Molecules* **2020**, *25*, 1259. [[CrossRef](#)] [[PubMed](#)]
23. Ogrodniczuk, M.; Motilall, A.; Fuanta, R. Shikimate kinase expression; towards screening of potential antibacterials from marine sponges. *FASEB J.* **2020**, *34*, 1. [[CrossRef](#)]
24. Sivaranjani, P.; Naik, V.U.; Madhulitha, N.R.; Kumar, K.S.; Chiranjeevi, P.; Alexander, S.P.; Umamaheswari, A. Design of novel antimycobacterial molecule targeting shikimate pathway of *Mycobacterium tuberculosis*. *Indian J. Pharm. Sci.* **2019**, *81*, 438–447. [[CrossRef](#)]
25. Pereira, J.H.; Vasconcelos, I.B.; Oliveira, J.S.; Caceres, R.A.; de Azevedo, W.F.; Basso, L.A.; Santos, D.S. Shikimate kinase: A potential target for development of novel antitubercular agents. *Curr. Drug Targets* **2007**, *8*, 459–468. [[CrossRef](#)]
26. Dias, M.V.B.; Ely, F.; Palma, M.S.; de Azevedo, W.F.; Basso, L.A.; Santos, D.S. Chorismate synthase: An attractive target for drug development against orphan diseases. *Curr. Drug Targets* **2007**, *8*, 437–444. [[CrossRef](#)]
27. Khanapur, M.; Alvala, M.; Prabhakar, M.; Shiva Kumar, K.; Edwin, R.K.; Sri Saranya, P.S.V.K.; Patel, R.K.; Bulusu, G.; Misra, P.; Pal, M. *Mycobacterium tuberculosis* chorismate mutase: A potential target for TB. *Bioorg. Med. Chem.* **2017**, *25*, 1725–1736. [[CrossRef](#)]
28. González-Bello, C. Specific chemical modification of bacterial type I dehydroquinase—Opportunities for drug discovery. *Future Med. Chem.* **2015**, *7*, 2371–2383. [[CrossRef](#)]
29. De Avila, M.B.; Bitencourt-Ferreira, G.; de Azevedo, W.F., Jr. Structural basis for inhibition of enoyl-[acyl carrier protein] reductase (InhA) from *Mycobacterium tuberculosis*. *Curr. Med. Chem.* **2020**, *27*, 745–759. [[CrossRef](#)]
30. Agoni, C.; Olotu, F.A.; Ramharack, P.; Soliman, M.E. Druggability and drug-likeness concepts in drug design: Are biomodelling and predictive tools having their say? *J. Mol. Model.* **2020**, *26*, 120. [[CrossRef](#)]
31. Barril, X. Druggability predictions: Methods, limitations, and applications. *WIREs Comput. Mol. Sci.* **2013**, *3*, 327–338. [[CrossRef](#)]
32. Macarron, R. Critical review of the role of HTS in drug discovery. *Drug Discov. Today* **2006**, *11*, 277–279. [[CrossRef](#)] [[PubMed](#)]

33. Michel, M.; Homan, E.J.; Wiita, E.; Pedersen, K.; Almlof, I.; Gustavsson, A.-L.; Lundback, T.; Helleday, T.; Berglund, U.W. In silico druggability assessment of the NUDIX hydrolase protein family as a workflow for target prioritization. *Front. Chem.* **2020**, *8*, 443. [[CrossRef](#)] [[PubMed](#)]
34. Hopkins, A.L.; Groom, C.R. Target analysis: A priori assessment of druggability. In *Proceedings of the Small Molecule—Protein Interactions*; Waldmann, H., Koppitz, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2003; pp. 11–17.
35. Hajduk, P.J.; Huth, J.R.; Fesik, S.W. Druggability indices for protein targets derived from NMR-based screening data. *J. Med. Chem.* **2005**, *48*, 2518–2525. [[CrossRef](#)]
36. Vukovic, S.; Huggins, D.J. Quantitative metrics for drug–target ligandability. *Drug Discov. Today* **2018**, *23*, 1258–1266. [[CrossRef](#)] [[PubMed](#)]
37. Gilson, M.K.; Liu, T.; Baitaluk, M.; Nicola, G.; Hwang, L.; Chong, J. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.* **2016**, *44*, D1045–D1053. [[CrossRef](#)]
38. Liu, T.; Lin, Y.; Wen, X.; Jorissen, R.N.; Gilson, M.K. BindingDB: A web-accessible database of experimentally determined protein–ligand binding affinities. *Nucleic Acids Res.* **2007**, *35*, D198–D201. [[CrossRef](#)]
39. Kozakov, D.; Grove, L.E.; Hall, D.R.; Bohnuud, T.; Mottarella, S.; Luo, L.; Xia, B.; Beglov, D.; Vajda, S. The FTMap Family of web servers for determining and characterizing ligand binding hot spots of proteins. *Nat. Protoc.* **2015**, *10*, 733–755. [[CrossRef](#)]
40. Volkamer, A.; Kuhn, D.; Rippmann, F.; Rarey, M. DoGSiteScorer: A web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics* **2012**, *28*, 2074–2075. [[CrossRef](#)]
41. *Schrödinger Release 2021-3: SiteMap*; Schrödinger, LLC.: New York, NY, USA, 2021.
42. Le Guilloux, V.; Schmidtke, P.; Tuffery, P. Fpocket: An open source platform for ligand pocket detection. *BMC Bioinform.* **2009**, *10*, 168. [[CrossRef](#)]
43. Hernandez, M.; Ghersi, D.; Sanchez, R. SITEHOUND-Web: A server for ligand binding site identification in protein structures. *Nucleic Acids Res.* **2009**, *37*, W413–W416. [[CrossRef](#)]
44. Liu, Q.; Li, L.; Xu, F. Systematic analysis and integrative discovery of active-site subpocket-specific dehydroquinase synthase inhibitors combating antibiotic-resistant *Staphylococcus aureus* infection. *J. Bioinform. Comput. Biol.* **2018**, *16*, 1850027. [[CrossRef](#)] [[PubMed](#)]
45. Wang, K.; Zhu, M.; Tang, Y.; Liu, J.; Yan, F.; Yu, Z.; Zhu, J. Integration of virtual screening and susceptibility test to discover active-site subpocket-specific biogenic inhibitors of helicobacter pylori shikimate dehydrogenase. *Int. Microbiol.* **2019**, *22*, 69–80. [[CrossRef](#)] [[PubMed](#)]
46. Gladki, A.; Kaczanowski, S.; Szczesny, P.; Zielenkiewicz, P. The evolutionary rate of antibacterial drug targets. *BMC Bioinform.* **2013**, *14*, 36. [[CrossRef](#)] [[PubMed](#)]
47. Zhao, H.; Gao, H.; Ji, K.; Yan, B.; Li, Q.; Mo, S.; Zheng, M.; Ou, Q.; Wu, B.; Li, N.; et al. Isolation and biochemical characterization of a metagenome-derived 3-deoxy-d-arabino-heptulosonate-7-phosphate synthase gene from subtropical marine mangrove wetland sediments. *Amb Express* **2019**, *9*, 19. [[CrossRef](#)] [[PubMed](#)]
48. Reiling, S.; Kelleher, A.; Matsumoto, M.M.; Robinson, G.; Asojo, O.A. Structure of type II dehydroquinase from *Pseudomonas aeruginosa*. *Acta Crystallogr. Sect. F-Struct. Biol. Commun.* **2014**, *70*, 1485–1491. [[CrossRef](#)]
49. Iqbal, N.; Kumar, M.; Sharma, P.; Yadav, S.P.; Kaur, P.; Sharma, S.; Singh, T.P. Binding studies and structure determination of the recombinantly produced type-II 3-dehydroquinase dehydratase from *Acinetobacter baumannii*. *Int. J. Biol. Macromol.* **2017**, *94*, 459–465. [[CrossRef](#)]
50. Florova, G.; Denoya, C.D.; Morgenstern, M.R.; Skinner, D.D.; Reynolds, K.A. Cloning, expression, and characterization of a type II 3-dehydroquinase dehydratase gene from *Streptomyces hygroscopicus*. *Arch. Biochem. Biophys.* **1998**, *350*, 298–306. [[CrossRef](#)]
51. Light, S.H.; Minasov, G.; Shuvalova, L.; Duban, M.-E.; Caffrey, M.; Anderson, W.F.; Lavie, A. Insights into the mechanism of type I dehydroquinase dehydratases from structures of reaction intermediates. *J. Biol. Chem.* **2011**, *286*, 3531–3539. [[CrossRef](#)]
52. Peek, J.; Christendat, D. The shikimate dehydrogenase family: Functional diversity within a conserved structural and mechanistic framework. *Arch. Biochem. Biophys.* **2015**, *566*, 85–99. [[CrossRef](#)]
53. Han, C.; Wang, L.; Yu, K.; Chen, L.; Hu, L.; Chen, K.; Jiang, H.; Shen, X. Biochemical characterization and inhibitor discovery of shikimate dehydrogenase from helicobacter pylori. *Febs J.* **2006**, *273*, 4682–4692. [[CrossRef](#)]
54. Michel, G.; Roszak, A.W.; Sauve, V.; Maclean, J.; Matte, A.; Coggins, J.R.; Cygler, M.; Laphorn, A.J. Structures of shikimate dehydrogenase AroE and its paralog YdiB—A common structural framework for different activities. *J. Biol. Chem.* **2003**, *278*, 19463–19472. [[CrossRef](#)] [[PubMed](#)]
55. Bagautdinov, B.; Kunishima, N. Crystal structures of shikimate dehydrogenase AroE from *Thermus thermophilus* HB8 and its cofactor and substrate complexes: Insights into the enzymatic mechanism. *J. Mol. Biol.* **2007**, *373*, 424–438. [[CrossRef](#)]
56. Peek, J.; Lee, J.; Hu, S.; Senisterra, G.; Christendat, D. Structural and mechanistic analysis of a novel class of shikimate dehydrogenases: Evidence for a conserved catalytic mechanism in the shikimate dehydrogenase family. *Biochemistry* **2011**, *50*, 8616–8627. [[CrossRef](#)] [[PubMed](#)]
57. Arcuri, H.A.; Borges, J.C.; Fonseca, I.O.; Pereira, J.H.; Ruggiero Neto, J.; Basso, L.A.; Santos, D.S.; de Azevedo, W.F. Structural studies of shikimate 5-dehydrogenase from *Mycobacterium tuberculosis*. *Proteins-Struct. Funct. Bioinforma.* **2008**, *72*, 720–730. [[CrossRef](#)] [[PubMed](#)]

58. Favela-Candia, A.; Tellez-Valencia, A.; Campos-Almazan, M.; Sierra-Campos, E.; Valdez-Solana, M.; Oria-Hernandez, J.; Castillo-Villanueva, A.; Najera, H.; Avitia-Dominguez, C. Biochemical, kinetic, and computational structural characterization of shikimate kinase from methicillin-resistant *Staphylococcus aureus*. *Mol. Biotechnol.* **2019**, *61*, 274–285. [[CrossRef](#)]
59. Liu, Z.; Zhu, Q.; Li, Y.; Yu, J.; Wang, W.; Tan, R.; Zhou, J.; Liao, H. Isolation and in silico characterization of a shikimate kinase from *Cassia obtusifolia*. *Acta Physiol. Plant.* **2015**, *37*, 85. [[CrossRef](#)]
60. Ahmad, S.; Baseer, S.; Navid, A.; Ahmad, F.; Azam, S.S. An integrated computational hierarchy for identification of potent inhibitors against shikimate kinase enzyme from *Shigella sonnei*, a major cause of global dysentery. *Gene Rep.* **2018**, *11*, 283–293. [[CrossRef](#)]
61. Rainio, M.J.; Ruuskanen, S.; Helander, M.; Saikkonen, K.; Saloniemi, I.; Puigbò, P. Adaptation of bacteria to glyphosate: A microevolutionary perspective of the enzyme 5-enolpyruvylshikimate-3-phosphate synthase. *Environ. Microbiol. Rep.* **2021**, *13*, 309–316. [[CrossRef](#)]
62. Leino, L.; Tall, T.; Helander, M.; Saloniemi, I.; Saikkonen, K.; Ruuskanen, S.; Puigbò, P. Classification of the glyphosate target enzyme (5-enolpyruvylshikimate-3-phosphate synthase) for assessing sensitivity of organisms to the herbicide. *J. Hazard. Mater.* **2021**, *408*, 124556. [[CrossRef](#)]
63. Tapas, S.; Kumar, A.; Dhindwal, S.; Preeti; Kumar, P. Structural analysis of chorismate synthase from *Plasmodium falciparum*: A novel target for antimalaria drug discovery. *Int. J. Biol. Macromol.* **2011**, *49*, 767–777. [[CrossRef](#)]
64. Neetu, N.; Sharma, M.; Mahto, J.K.; Kumar, P. Biophysical and in-silico studies of phytochemicals targeting chorismate synthase from drug-resistant *Moraxella catarrhalis*. *Protein J.* **2020**, *39*, 449–460. [[CrossRef](#)]
65. Kukol, A.; Hughes, D.J. Large-scale analysis of influenza A virus nucleoprotein sequence conservation reveals potential drug-target sites. *Virology* **2014**, *454–455*, 40–47. [[CrossRef](#)] [[PubMed](#)]
66. Trigueiro-Louro, J.M.; Correia, V.; Santos, L.A.; Guedes, R.C.; Brito, R.M.M.; Rebelo-de-Andrade, H. To hit or not to hit: Large-scale sequence analysis and structure characterization of influenza A NS1 unlocks new antiviral target potential. *Virology* **2019**, *535*, 297–307. [[CrossRef](#)] [[PubMed](#)]
67. Chen, J.W.; Romero, P.; Uversky, V.N.; Dunker, A.K. Conservation of intrinsic disorder in protein domains and families: II. functions of conserved disorder. *J. Proteome Res.* **2006**, *5*, 888–898. [[CrossRef](#)] [[PubMed](#)]
68. Chong, B.; Li, M.; Li, T.; Yu, M.; Zhang, Y.; Liu, Z. Conservation of potentially druggable cavities in intrinsically disordered proteins. *ACS Omega* **2018**, *3*, 15643–15652. [[CrossRef](#)] [[PubMed](#)]
69. Landau, M.; Mayrose, I.; Rosenberg, Y.; Glaser, F.; Martz, E.; Pupko, T.; Ben-Tal, N. ConSurf 2005: The projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **2005**, *33*, W299–W302. [[CrossRef](#)]
70. Jiao, W.; Blackmore, N.J.; Nazmi, A.R.; Parker, E.J. Quaternary structure is an essential component that contributes to the sophisticated allosteric regulation mechanism in a key enzyme from *Mycobacterium tuberculosis*. *PLoS ONE* **2017**, *12*, e0180052. [[CrossRef](#)]
71. Jensen, R.A.; Xie, G.; Calhoun, D.H.; Bonner, C.A. The correct phylogenetic relationship of KdsA (3-deoxy-D-manno-octulosonate 8-phosphate synthase) with one of two independently evolved classes of AroA (3-deoxy-D-arabino-heptulosonate 7-phosphate synthase). *J. Mol. Evol.* **2002**, *54*, 416–423. [[CrossRef](#)]
72. Webby, C.J.; Baker, H.M.; Lott, J.S.; Baker, E.N.; Parker, E.J. The structure of 3-deoxy-d-arabino-heptulosonate 7-phosphate synthase from *Mycobacterium tuberculosis* reveals a common catalytic scaffold and ancestry for type I and type II enzymes. *J. Mol. Biol.* **2005**, *354*, 927–939. [[CrossRef](#)]
73. Light, S.H.; Anderson, W.F. The diversity of allosteric controls at the gateway to aromatic amino acid biosynthesis. *Protein Sci.* **2013**, *22*, 395–404. [[CrossRef](#)] [[PubMed](#)]
74. Subramaniam, P.S.; Xie, G.; Xia, T.H.; Jensen, R.A. Substrate ambiguity of 3-deoxy-D-manno-octulosonate 8-phosphate synthase from *Neisseria gonorrhoeae* in the context of its membership in a protein family containing a subset of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthases. *J. Bacteriol.* **1998**, *180*, 119–127. [[CrossRef](#)] [[PubMed](#)]
75. Sterritt, O.W.; Kessans, S.A.; Jameson, G.B.; Parker, E.J. A pseudoisomeric type II DAH7PS enzyme from *Pseudomonas aeruginosa*: Alternative evolutionary strategies to control shikimate pathway flux. *Biochemistry* **2018**, *57*, 2667–2678. [[CrossRef](#)] [[PubMed](#)]
76. Harris, J.M.; GonzalezBello, C.; Kleanthous, C.; Hawkins, A.R.; Coggins, J.R.; Abell, C. Evidence from kinetic isotope studies for an enolate intermediate in the mechanism of type II dehydroquinases. *Biochem. J.* **1996**, *319*, 333–336. [[CrossRef](#)]
77. TRomanowski, M.J.; Burley, S.K. Crystal structure of the *Escherichia coli* shikimate kinase I (AroK) that confers sensitivity to mecillinam. *Proteins-Struct. Funct. Genet.* **2002**, *47*, 558–562. [[CrossRef](#)]
78. Kortemme, T.; Baker, D. A Simple physical model for binding energy hot spots in protein–protein complexes. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14116–14121. [[CrossRef](#)]
79. DeLano, W.L. Unraveling hot spots in binding interfaces: Progress and challenges. *Curr. Opin. Struct. Biol.* **2002**, *12*, 14–20. [[CrossRef](#)]
80. Halgren, T.A. Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inf. Model.* **2009**, *49*, 377–389. [[CrossRef](#)]
81. Sharir-Ivry, A.; Xia, Y. Quantifying evolutionary importance of protein sites: A tale of two measures. *PLoS Genet.* **2021**, *17*, e1009476. [[CrossRef](#)]

82. Ma, B.; Elkayam, T.; Wolfson, H.; Nussinov, R. Protein–protein interactions: Structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 5772–5777. [[CrossRef](#)]
83. Slodkiewicz, G.; Goldman, N. Integrated structural and evolutionary analysis reveals common mechanisms underlying adaptive evolution in mammals. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 5977–5986. [[CrossRef](#)] [[PubMed](#)]
84. Bornemann, S.; Ramjee, M.; Balasubramanian, S.; Abell, C.; Coggins, J.; Lowe, D.; Thorneley, R. *Escherichia coli* chorismate synthase catalyzes the conversion of (6S)-6-fluoro-5-enolpyruvylshikimate-3-phosphate to 6-fluorochorismate—Implications for the enzyme mechanism and the antimicrobial action of (6S)-6-fluoroshikimate. *J. Biol. Chem.* **1995**, *270*, 22811–22815. [[CrossRef](#)] [[PubMed](#)]
85. Hasan, N.; Nester, E. Dehydroquinase synthase in bacillus-subtilis—Enzyme associated with chorismate synthase and flavin reductase. *J. Biol. Chem.* **1978**, *253*, 4999–5004. [[CrossRef](#)]
86. Bornemann, S.; Lowe, D.J.; Thorneley, R.N.F. The transient kinetics of *Escherichia coli* chorismate synthase: Substrate, consumption, product formation, phosphate dissociation, and characterization of a flavin intermediate. *Biochemistry* **1996**, *35*, 9907–9916. [[CrossRef](#)] [[PubMed](#)]
87. Macheroux, P.; Schonbrunn, E.; Svergun, D.I.; Volkov, V.V.; Koch, M.H.J.; Bornemann, S.; Thorneley, R.N.F. Evidence for a major structural change in *Escherichia coli* chorismate synthase induced by flavin and substrate binding. *Biochem. J.* **1998**, *335*, 319–327. [[CrossRef](#)]
88. Maclean, J.; Ali, S. The structure of chorismate synthase reveals a novel flavin binding to a unique chemical reaction. *Structure* **2003**, *11*, 1499–1511. [[CrossRef](#)]
89. Capra, J.A.; Laskowski, R.A.; Thornton, J.M.; Singh, M.; Funkhouser, T.A. Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure. *PLoS Comput. Biol.* **2009**, *5*, e1000585. [[CrossRef](#)]
90. Christopoulos, A. Allosteric binding sites on cell-surface receptors: Novel targets for drug discovery. *Nat. Rev. Drug Discov.* **2002**, *1*, 198–210. [[CrossRef](#)]
91. Wenthur, C.J.; Gentry, P.R.; Mathews, T.P.; Lindsley, C.W. Drugs for allosteric sites on receptors. *Annu. Rev. Pharmacol. Toxicol.* **2014**, *54*, 165–184. [[CrossRef](#)]
92. Mishra, A.; Mamidi, A.S.; Rajmani, R.S.; Ray, A.; Roy, R.; Surolia, A. An allosteric inhibitor of *Mycobacterium tuberculosis* ArgJ: Implications to a novel combinatorial therapy. *EMBO Mol. Med.* **2018**, *10*, e8038. [[CrossRef](#)]
93. Wellington, S.; Nag, P.P.; Michalska, K.; Johnston, S.E.; Jedrzejczak, R.P.; Kaushik, V.K.; Clatworthy, A.E.; Siddiqi, N.; McCarren, P.; Bajrami, B.; et al. A small-molecule allosteric inhibitor of *Mycobacterium tuberculosis* tryptophan synthase. *Nat. Chem. Biol.* **2017**, *13*, 943–950. [[CrossRef](#)] [[PubMed](#)]
94. Ivetac, A.; Andrew McCammon, J. Mapping the druggable allosteric space of G-protein coupled receptors: A fragment-based molecular dynamics approach. *Chem. Biol. Drug Des.* **2010**, *76*, 201–217. [[CrossRef](#)] [[PubMed](#)]
95. Robertson, J.C.; Hurley, N.C.; Tortorici, M.; Cioffani, G.; Borrello, M.T.; Vellore, N.A.; Ganesan, A.; Mattevi, A.; Baron, R. Expanding the druggable space of the LSD1/CoREST epigenetic target: New potential binding regions for drug-like molecules, peptides, protein partners, and chromatin. *PLoS Comput. Biol.* **2013**, *9*, e1003158. [[CrossRef](#)] [[PubMed](#)]
96. Vittorio, S.; Seidel, T.; Garon, A.; Gitto, R.; Langer, T.; De Luca, L. In silico identification of potential druggable binding sites on CIN85 SH3 domain. *Int. J. Mol. Sci.* **2021**, *22*, 534. [[CrossRef](#)]
97. Borota, A.; Halip, L.; Curpan, R.; Bora, A.; Avram, S.; Mracec, M. Structure- and ligand- based studies to gain insight into the pharmacological implications of histamine H-3 receptor. *Struct. Chem.* **2021**, *32*, 1141–1149. [[CrossRef](#)]
98. Veeravarapu, H.; Malkhed, V.; Mustyala, K.K.; Vadija, R.; Malikanti, R.; Vuruputuri, U.; Muthyala, M.K.K. Structure-based drug design, synthesis and screening of MmaA1 inhibitors as novel anti-TB agents. *Mol. Divers.* **2021**, *25*, 351–366. [[CrossRef](#)] [[PubMed](#)]
99. Muddagoni, N.; Bathula, R.; Dasari, M.; Potlapally, S.R. Homology modeling, virtual screening, prime-MMGBSA, AutoDock-identification of inhibitors of FGR protein. *Biointerface Res. Appl. Chem.* **2021**, *11*, 11088–11103. [[CrossRef](#)]
100. Roky, M.; Trent, J.O.; Demuth, D.R. Identification of functional domains of the minor fimbrial antigen involved in the interaction of *Porphyromonas gingivalis* with oral streptococci. *Mol. Oral Microbiol.* **2020**, *35*, 66–77. [[CrossRef](#)]
101. Bathula, R.; Lanka, G.; Muddagoni, N.; Dasari, M.; Nakkala, S.; Bhargavi, M.; Somadi, G.; Sivan, S.K.; Potlapally, S.R. Identification of potential aurora kinase-C protein inhibitors: An amalgamation of energy minimization, virtual screening, prime MMGBSA and AutoDock. *J. Biomol. Struct. Dyn.* **2020**, *38*, 2314–2325. [[CrossRef](#)]
102. Zhou, W.; Lu, S.; Zhang, Y.; Yin, L.; Zhu, L.; Zhao, J.; Lu, T.; Chen, Y.; Liu, H. In silico design of new B-raf kinase type-II inhibitors through combined molecular modeling studies. *Lett. Drug Des. Discov.* **2019**, *16*, 570–583. [[CrossRef](#)]
103. Verma, P.; Tiwari, V. Targeting outer membrane protein component AdeC for the discovery of efflux pump inhibitor against AdeABC efflux pump of multidrug resistant *Acinetobacter baumannii*. *Cell Biochem. Biophys.* **2018**, *76*, 391–400. [[CrossRef](#)]
104. Usman, M.M.S.; Bharbhuiya, T.K.; Mondal, S.; Rani, S.; Kyal, C.; Kumari, R. Combined protein and ligand based physicochemical aspects of molecular recognition for the discovery of CDK9 inhibitor. *Gene Rep.* **2018**, *13*, 212–219. [[CrossRef](#)]
105. Bowkett, D.; Talon, R.; Tallant, C.; Schofield, C.; von Delft, F.; Knapp, S.; Bruton, G.; Brennan, P.E. Identifying small-molecule binding sites for epigenetic proteins at domain-domain interfaces. *Chemmedchem* **2018**, *13*, 1051–1057. [[CrossRef](#)] [[PubMed](#)]
106. Adeniji, E.A.; Olotu, F.A.; Soliman, M.E.S. Exploring the lapse in druggability: Sequence analysis, structural dynamics and binding site characterization of K-Ras(G12C) variant, a feasible oncotherapeutics target. *Anticancer Agents Med. Chem.* **2018**, *18*, 1540–1550. [[CrossRef](#)]

107. Olotu, F.A.; Omolabi, K.F.; Soliman, M.E.S. Leaving no stone unturned: Allosteric targeting of SARS-CoV-2 spike protein at putative druggable sites disrupts human angiotensin-converting enzyme interactions at the receptor binding domain. *Inform. Med. Unlocked* **2020**, *21*, 100451. [CrossRef] [PubMed]
108. David-Eden, H.; Mankin, A.S.; Mandel-Gutfreund, Y. Structural signatures of antibiotic binding sites on the ribosome. *Nucleic Acids Res.* **2010**, *38*, 5982–5994. [CrossRef] [PubMed]
109. Ayyildiz, M.; Celiker, S.; Ozhelvacı, F.; Akten, E.D. Identification of alternative allosteric sites in glycolytic enzymes for potential use as species-specific drug targets. *Front. Mol. Biosci.* **2020**, *7*, 88. [CrossRef]
110. Sheik Amamuddy, O.; Veldman, W.; Manyumwa, C.; Khairallah, A.; Agajanian, S.; Oluyemi, O.; Verkhivker, G.M.; Tastan Bishop, Ö. Integrated computational approaches and tools for allosteric drug discovery. *Int. J. Mol. Sci.* **2020**, *21*, 847. [CrossRef]
111. Perola, E.; Herman, L.; Weiss, J. Development of a rule-based method for the assessment of protein druggability. *J. Chem. Inf. Model.* **2012**, *52*, 1027–1038. [CrossRef]
112. Hajduk, P.J.; Huth, J.R.; Tse, C. Predicting protein druggability. *Drug Discov. Today* **2005**, *10*, 1675–1682. [CrossRef]
113. Gao, M.; Skolnick, J. A Comprehensive survey of small-molecule binding pockets in proteins. *PLoS Comput. Biol.* **2013**, *9*, e1003302. [CrossRef] [PubMed]
114. Smith, R.D.; Lu, J.; Carlson, H.A. Are there physicochemical differences between allosteric and competitive ligands? *PLoS Comput. Biol.* **2017**, *13*, e1005813. [CrossRef] [PubMed]
115. Lu, S.; He, X.; Ni, D.; Zhang, J. Allosteric modulator discovery: From serendipity to structure-based design. *J. Med. Chem.* **2019**, *62*, 6405–6421. [CrossRef] [PubMed]
116. Chen, J.; Yang, H.; Zhu, L.; Wu, Z.; Li, W.; Tang, Y.; Liu, G. In silico prediction of human renal clearance of compounds using quantitative structure-pharmacokinetic relationship models. *Chem. Res. Toxicol.* **2020**, *33*, 640–650. [CrossRef] [PubMed]
117. Cheng, A.C.; Coleman, R.G.; Smyth, K.T.; Cao, Q.; Soulard, P.; Caffrey, D.R.; Salzberg, A.C.; Huang, E.S. Structure-based maximal affinity model predicts small-molecule druggability. *Nat. Biotechnol.* **2007**, *25*, 71–75. [CrossRef]
118. Berthold, M.R.; Cebon, N.; Dill, F.; Gabriel, T.R.; Kötter, T.; Meinel, T.; Ohl, P.; Thiel, K.; Wiswedel, B. KNIME—The Konstanz information miner: Version 2.0 and beyond. *ACM SIGKDD Explor. Newsl.* **2009**, *11*, 26–31. [CrossRef]
119. Ashkenazy, H.; Erez, E.; Martz, E.; Pupko, T.; Ben-Tal, N. ConSurf 2010: Calculating Evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **2010**, *38*, W529–W533. [CrossRef]
120. Celniker, G.; Nimrod, G.; Ashkenazy, H.; Glaser, F.; Martz, E.; Mayrose, I.; Pupko, T.; Ben-Tal, N. ConSurf: Using evolutionary data to raise testable hypotheses about protein function. *Isr. J. Chem.* **2013**, *53*, 199–206. [CrossRef]
121. Ashkenazy, H.; Abadi, S.; Martz, E.; Chay, O.; Mayrose, I.; Pupko, T.; Ben-Tal, N. ConSurf 2016: An Improved Methodology to Estimate and Visualize Evolutionary Conservation in Macromolecules. *Nucleic Acids Res.* **2016**, *44*, W344–W350. Available online: <https://academic.oup.com/nar/article/44/W1/W344/2499373> (accessed on 21 June 2021). [CrossRef]
122. *The PyMOL Molecular Graphics System, Version 2.3.3*; Schrödinger, LLC.: New York, NY, USA, 2019.
123. Gonçalves, W.R.S.; Gonçalves-Almeida, V.M.; Arruda, A.L.; Meira, W., Jr.; da Silveira, C.H.; Pires, D.E.V.; de Melo-Minardi, R.C. PDBest: A user-friendly platform for manipulating and enhancing protein structures. *Bioinformatics* **2015**, *31*, 2894–2896. [CrossRef]
124. KNIME: The Konstanz Information Miner. SpringerLink. Available online: https://link.springer.com/chapter/10.1007/978-3-54-0-78246-9_38 (accessed on 21 June 2021).
125. *Schrödinger Release 2021-3: Maestro*; Schrödinger, LLC.: New York, NY, USA, 2021.
126. UniProt consortium UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49*, D480–D489. [CrossRef] [PubMed]
127. KEGG: Kyoto Encyclopedia of Genes and Genomes. Available online: <http://www.kegg.jp/> (accessed on 14 October 2021).
128. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef] [PubMed]
129. Edgar, R.C. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.* **2004**, *5*, 113. [CrossRef] [PubMed]
130. Henikoff, S.; Henikoff, J.G. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 10915–10919. [CrossRef]
131. Mayrose, I.; Graur, D.; Ben-Tal, N.; Pupko, T. Comparison of site-specific rate-inference methods for protein sequences: Empirical bayesian methods are superior. *Mol. Biol. Evol.* **2004**, *21*, 1781–1791. [CrossRef]
132. Pupko, T.; Graur, D. Fast computation of maximum likelihood trees by numerical approximation of amino acid replacement probabilities. *Comput. Stat. Data Anal.* **2002**, *40*, 285–291. [CrossRef]
133. Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A.P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L.J.; Cibrián-Uhalte, E.; et al. The ChEMBL database in 2017. *Nucleic Acids Res.* **2017**, *45*, D945–D954. [CrossRef]
134. Shelley, J.C.; Cholleti, A.; Frye, L.L.; Greenwood, J.R.; Timlin, M.R.; Uchimaya, M. Epik: A software program for PK(a) prediction and protonation state generation for drug-like molecules. *J. Comput. Aided Mol. Des.* **2007**, *21*, 681–691. [CrossRef]
135. *Schrödinger Release 2022-1: Epik*; Schrödinger, LLC.: New York, NY, USA, 2021.
136. Waskom, M.L. Seaborn: Statistical data visualization. *J. Open Source Softw.* **2021**, *6*, 3021. [CrossRef]
137. Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [CrossRef]

138. Kluyver, T.; Ragan-Kelley, B.; Pérez, F.; Granger, B.; Bussonnier, M.; Frederic, J.; Kelley, K.; Hamrick, J.; Grout, J.; Corlay, S.; et al. Jupyter notebooks—A publishing format for reproducible computational workflows. *Position. Power Acad. Publ. Play. Agents Agendas* **2016**, *87–90*. [[CrossRef](#)]
139. Blanco, B.; Prado, V.; Lence, E.; Otero, J.M.; Garcia-Doval, C.; van Raaij, M.J.; Llamas-Saiz, A.L.; Lamb, H.; Hawkins, A.R.; Gonzalez-Bello, C. *Mycobacterium tuberculosis* Shikimate Kinase Inhibitors: Design and Simulation Studies of the Catalytic Turnover. *J. Am. Chem. Soc.* **2013**, *135*, 12366–12376. [[CrossRef](#)]
140. Balachandran, N.; Heimhalt, M.; Liuni, P.; To, F.; Wilson, D.J.; Junop, M.S.; Berti, P.J. Potent Inhibition of 3-Deoxy-D-Arabinose-7-Phosphate (DAHP) Synthase by DAHP Oxime, a Phosphate Group Mimic. *Biochemistry* **2016**, *55*, 6617–6629. [[CrossRef](#)] [[PubMed](#)]
141. Walker, S.R.; Jiao, W.; Parker, E.J. Synthesis and Evaluation of Dual Site Inhibitors of 3-Deoxy-D-Arabinose-7-Phosphate Synthase. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 5092–5097. [[CrossRef](#)] [[PubMed](#)]
142. Peterson, M.L.; Corey, S.D.; Font, J.L.; Walker, M.C.; Sikorski, J.A. New Simplified Inhibitors of EPSP Synthase: The Importance of Ring Size for Recognition at the Shikimate 3-Phosphate Site. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 2853–2858. [[CrossRef](#)]
143. Zhu, N.; Wang, X.; Li, D.; Lin, Y.; You, X.; Jiang, J.; Xu, Y.; Jiang, W.; Si, S. IMB-T130 Targets 3-Dehydroquinate Synthase and Inhibits *Mycobacterium tuberculosis*. *Sci. Rep.* **2018**, *8*, 17439. [[CrossRef](#)]
144. Montchamp, J.L.; Frost, J.W. Cyclohexenyl and Cyclohexylidene Inhibitors of 3-Dehydroquinate Synthase: Active Site Interactions Relevant to Enzyme Mechanism and Inhibitor Design. *J. Am. Chem. Soc.* **1997**, *119*, 7645–7653. [[CrossRef](#)]
145. Prado, V.; Lence, E.; Maneiro, M.; Vazquez-Ucha, J.C.; Beceiro, A.; Thompson, P.; Hawkins, A.R.; Gonzalez-Bello, C. Targeting the Motion of Shikimate Kinase: Development of Competitive Inhibitors That Stabilize an Inactive Open Conformation of the Enzyme. *J. Med. Chem.* **2016**, *59*, 5471–5487. [[CrossRef](#)]
146. Gonzalez-Bello, C.; Manthey, M.K.; Harris, J.H.; Hawkins, A.R.; Coggins, J.R.; Abell, C. Synthesis of 2-Bromo- and 2-Fluoro-3-Dehydroshikimic Acids and 2-Bromo- and 2-Fluoroshikimic Acids Using Synthetic and Enzymatic Approaches. *J. Org. Chem.* **1998**, *63*, 1591–1597. [[CrossRef](#)]
147. Lence, E.; Tizon, L.; Otero, J.M.; Peon, A.; Prazeres, V.F.V.; Llamas-Saiz, A.L.; Fox, G.C.; van Raaij, M.J.; Lamb, H.; Hawkins, A.R.; et al. Mechanistic Basis of the Inhibition of Type II Dehydroquinase by (2S)- and (2R)-2-Benzyl-3-Dehydroquinic Acids. *ACS Chem. Biol.* **2013**, *8*, 568–577. [[CrossRef](#)]
148. Thomas, M.G.; Lawson, C.; Allanson, N.M.; Leslie, B.W.; Bottomley, J.R.; McBride, A.; Olusanya, O.A. A Series of 2(Z)-2-Benzylidene-6,7-Dihydroxybenzofuran-3[2H]-Ones as Inhibitors of Chorismate Synthase. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 423–426. [[CrossRef](#)]
149. Peek, J.; Shi, T.; Christendat, D. Identification of Novel Polyphenolic Inhibitors of Shikimate Dehydrogenase (AroE). *J. Biomol. Screen.* **2014**, *19*, 1090–1098. [[CrossRef](#)] [[PubMed](#)]