Binding site comparisons for target centered drug discovery

Abstract

Introduction The success of comparison of binding sites in drug discovery is based on the recognized fact that many different proteins have similar binding sites. Binding site comparison has found many uses in drug development, and has the potential to dramatically cut the cost and shorten the time necessary for the development of new drugs.

Areas covered We review recent methods for comparing protein binding sites and their use in drug repurposing and polypharmacology. We then examine emerging fields, such as the use of binding site comparison in precision medicine, prediction of structured waters, search for targets of natural compounds, for use in development of protein-based drugs by loop modeling and for comparison of RNA binding sites.

Expert opinion This approach has produced many interesting results in drug development, but relatively little work has been done on protein-protein interaction sites, which are particularly relevant in view of the success of biological drugs. Therefore, growth of protein loop modeling for modulating biological drugs can be anticipated. The fusion of currently distinct methods for comparison of RNA and protein binding sites into a single comprehensive approach could allow the search for new selective ribosomal antibiotics and initiate pharmaceutical research into other nucleoproteins.

Keywords: Binding site comparison, drug repurposing, polypharmacology, precision medicine, conserved water, natural products, biological drugs, RNA motifs.

Article highlights

* Binding site comparison allows drug repurposing, an approach that is increasingly gaining use in treatment of rare and neglected diseases, as it allows significantly faster development than conventional drug development.

* Binding site comparison allows the design of polypharmacological drugs that could improve the treatment of cancer and neurodegenerative diseases.

* Binding site comparison enables the proteome-wide prediction of binding sites and ligands of proteins, and allows the determination of the effects of sequence variants on binding of drugs and development of diseases.

* Binding site comparison permits the identification of targets for natural products.

* Binding site comparison facilitates the prediction of new conserved structured waters that are important for the activity of kinases.

* Binding site comparison enables protein loop modeling in the development of biological drugs and allows prediction of RNA functions and the development of selective ribosome-targeting antibiotics.

1. Introduction

Successful drug development currently often takes longer than a decade and is becoming increasingly expensive. The cost of development of a single drug is currently well over $1bn. In order to reduce the cost of developing new drugs, it is important to recognize that although a drug was developed to bind to only one target, most of the drugs currently in use bind to multiple proteins. The pharmaceutical industry would benefit immensely from approaches that can predict or explain such multi-target drug binding. Binding site comparison is such an approach and has been used successfully in various drug development programs.

For some five decades, drug design has been based on matching a chemical structure with a specific binding site in a protein. Comparison of binding sites, a relatively new approach to drug development, inverts this problem and allows searching for binding sites in any protein that match a given chemical structure. The binding site comparison approach comprises various computer methods that enable the detection of similarities between proteins irrespective of sequence and fold similarities [1–12]. These methods are based on the fact that binding sites on proteins are more evolutionarily conserved than the rest of the protein structure. Common to the methods is that they compare structures of two binding sites with each other at one time, resulting in a computed degree of similarity of the two compared binding sites and in their three-dimensional superimposition. Pre-knowledge of one or both binding sites is not necessary; some methods compare the selected binding site against the entire protein structure [1,5], and some compare whole proteins against whole proteins [2,3]. Unlike global comparisons that compare protein backbones, these methods allow one to find locally-restricted similar interactions and structural patterns in proteins that are most often present within drug binding sites.

Homology docking is an extension of the binding site comparison approach and is based on the fact that similar binding sites bind similar ligands [13]. Thus, based on the superposition of their similar binding sites, ligands can be transferred between proteins [14,15]. Homology docking

enables prediction of the protein's ligands and binding sites, wherein a binding site is defined as the space around the protein occupied by transferred ligands of a particular type. Binding sites for entire proteomes have been computed and deposited in various databases [16–19].

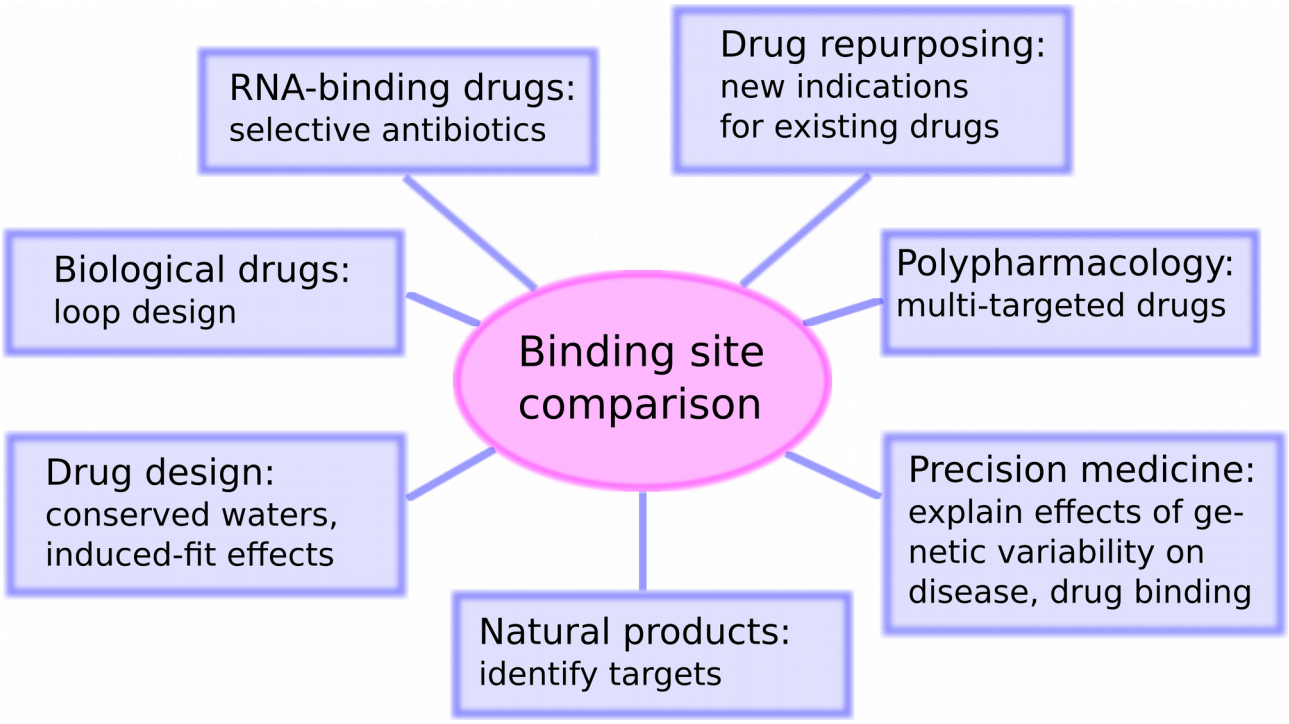

Figure 1. Use of binding site comparison in drug discovery.

The pharmaceutical community accepted binding site comparison mainly due to its use in drug repurposing [20]. A well-known example concerns the nonsteroidal anti-inflammatory drugs, celecoxib and valdecoxib, COX-2 inhibitors that were found to act at the nanomolar level as carbonic anhydrase II inhibitors. Based on findings from binding site comparison, these two drugs were suggested for the treatment of glaucoma and cancer, and were repurposed for these new indications [21]. The use of binding site comparisons in the design of polypharmacological drugs, those with various pharmacological behavior, is also prominent. A successful example of this is the

HIV protease inhibitor nelfinavir that was found to act polypharmacologically as an anticancer agent by inhibiting multiple kinases [22].

This article reviews the recent methods of binding site comparison and their use in drug discovery and development (Figure 1). In contrast to previous reviews [20,23], the focus is on new applications that represent promising new paths for the future development of the field and new methods of drug discovery.

2. Recent advances in binding site comparison

The first binding site comparison methods were developed over fifteen years ago, giving researchers insights into previously overlooked structural similarities and differences in protein binding sites [7]. Unlike the global alignment, comparison of binding sites reveals locally preserved patterns of amino acid residues, including those in unrelated proteins possessing different sequences and folds.

Until recently, comparison of binding sites was hindered by the lack of powerful computing resources, but in the last few years, massively parallel systems, such as the graphics processing units (GPU), allowing for much faster binding site comparison methods, have appeared on the scene. Thus, the new SEGA method implemented on the GPU [24] leads to qualitatively better comparison of binding sites and significantly shorter execution times compared to the earlier corresponding CPU version of the method. An achieved acceleration of up to 1,500 times opens the possibility, so far unexplored of rapid proteome-wide binding site comparison.

Comparison of protein-protein binding sites is less frequently used than comparison of protein-small molecule binding sites. This is likely due to the finding that protein-protein binding sites are evolutionarily less conserved than binding sites for small molecules [25]. Nevertheless, methods such as PRISM [14] and ProBiS [15] allow for comparison of protein-protein binding sites and prediction and modeling of new protein-protein interactions based on the homology docking approach.

Binding site comparison methods treat proteins as rigid structures, which creates an issue in homology docking [14,15]. Due to the rigidity of the compared protein structures, the predicted protein-ligand complexes may be unrealistic due to steric clashes. In order to overcome this issue the ProBiS-CHARMMing web interface [26] was developed to allow flexible homology docking of ligands such as proteins, nucleic acids, small molecules and ions. Energy minimization is used to remove steric clashes between the predicted ligands and proteins, thus correcting the complex structures.

In any research field, uniform standards for comparing newly developed methods and independent tests for their evaluation are very important. Standardized test sets have already been used successfully in molecular docking [27], and have led to more effective development in this field. On the contrary, a wide and diverse set of tests is used to evaluate binding site comparison methods, which makes it difficult to draw conclusions concerning the success of different methods. Recently, ProSPECCTs, a collection of test binding sites [28] was prepared for this purpose. It can be used to evaluate new algorithms for binding site comparison. In addition, it also helps to answer the question as to which method to use in any particular case. Indeed, for comparison of very similar binding sites, such as the kinase binding sites, we need different methods than comparison of weakly similar binding sites in evolutionarily distant proteins that are considered in drug repurposing and polypharmacology.

In the subsequent sections we describe the use of different binding site comparison methods in various pharmaceutical applications.

3. Drug repurposing

The drug repurposing approach is becoming increasingly relevant for the development of drugs to treat rare and neglected diseases, diseases that only outbreak periodically, as well as for bacterial diseases. Its use in rare immunological disorders has been reviewed elsewhere [29] and will not be discussed here. Although rare and neglected diseases together pose a significant health problem,

large pharmaceutical companies tend not to develop new drugs for them, due to the small potential markets. The main characteristic of the drug repurposing is that it allows the repositioning of drugs registered and approved by the US Food and Drug Administration (FDA) drugs and drug candidates to new indications [30]. Since these molecules have already, at least partly, passed costly clinical trials, the approach promises faster and cheaper development than classical *de novo* development. One of the methods in drug repurposing is homology docking based on binding site comparison, which is illustrated by recent interesting cases in the following Sections.

3.1. Diseases that escalate periodically

To identify new drug candidates to treat Ebola, a viral hemorrhagic fever, an integrated pharmacological pipeline was developed that complemented binding site comparison with molecular docking and molecular dynamics simulations [31]. From the 1,766 FDA-approved drugs and 259 drug candidates from DrugBank [32] this pipeline identified molecules that could inhibit Ebola replication and virulence. The study focused on two main targets, the RNA-directed RNA polymerase, and the multifunctional secondary matrix protein VP24. The assumed binding site on VP24 was compared with binding sites in 40,491 biological assemblies from the Protein Data Bank (PDB) [33], wherein amino acids in the proximity of the co-crystallized ligands were considered as binding sites. It was found that VP24 and HIV protease share similar binding sites and was suggested that indinavir, an HIV protease inhibitor, could probably reduce Ebola virulence effectively. This finding was additionally supported by molecular docking of the FDA-approved drugs, which ranked indinavir among the top binders of the VP24 protein. Similarly, it was found that the antifungal agent sinefungin likely inhibits RNA-directed RNA polymerase. Binding site comparison in this case proved to be very useful for drug repurposing in combination with other complementary methods.

3.2. Rare and neglected diseases

Binding site comparison in combination with ligand- and structure-based virtual screening has been used to repurpose drugs to proteins involved in the pathogenesis of rare diseases [23]. Drugs from the DrugBank [32] were transferred to the proteins from the Orphanet database [34], with the result that 31,142 protein-drug complexes were associated with 980 orphan diseases, diseases that affect <200,000 people in the US. It was found that steroidal aromatase inhibitors, a class of drugs used to treat breast cancer, could also be used for the treatment of a rare inherited metabolic Niemann-Pick type C disease.

In another study, a target-hopping approach based on binding site comparison was used to repurpose drugs for the Chagas disease, widespread in Latin America and caused by the *Trypanosoma cruzi* parasite [35]. Here, homology docking was used with an additional approach to measure the degree of similarity between the transferred ligand and the native ligand of the target protein which increased the likelihood that the predicted ligand will form similar interactions as the native ligand [36]. The approach was validated with the drugs that are already used to treat the Chagas disease and then, new putative drugs were predicted. Foscarnet, a known antiviral drug appeared particularly promising and was predicted to target farnesyl pyrophosphate synthase in *T. cruzi*.

3.3. Bacterial diseases

Novel scaffolds open new opportunities to medicinal chemists for the development of inhibitors with properties different from those already known [37]. Homology docking was used to extend the scaffold range of inhibitors of pathway II for the biosynthesis of mycobacterial fatty acids, which is an attractive target for the development of selective antimycobacterial agents [38]. The study focused on InhA, the key enzyme on this pathway, and an NADH-dependent enzyme enoyl-acyl carrier protein reductase targeted by isoniazid, the original antimycobacterial drug [39]. Almost 600 ligands were transferred to InhA, and eight were selected for experimental evaluation [38]. Those

selected originated from proteins with low sequence identities (~30%) to the InhA enzyme and with different scaffolds than the known InhA inhibitors. Three active compounds were confirmed, 1-(3,4-dichlorobenzyl)-5,6-dimethyl-1H-benzo[d]imidazole being the most active with an IC_{50} value of 10 ± 2 μM. The three new InhA inhibitors had new scaffolds previously untested on InhA, and represented good starting points for further optimization. The advantage of binding site comparison is that it allows the discovery from related research fields of new compounds, which were previously not considered as antimycotic agents.

4. Polypharmacology

The fact that most drugs are non-selective is a problem in pharmacology and gives rise to unwanted off-target events in patients [40]. However, many diseases are the result of complex biological processes and to overcome the robustness of such complex diseases, it is necessary to disrupt the functioning of several receptors. Polypharmacology aims to design drugs in such a way that they bind to multiple targets, whose simultaneous inhibition is synergistic. Polypharmacology shows potential for the development of new drugs for polygenic diseases, such as neurodegenerative diseases and cancers, where multi-targeting can reduce the onset of resistance. Binding site comparison plays a key role in polypharmacology due to its ability to predict new targets [20].

4.1. Kinases as polypharmacological targets

Following the approval in 2001 of the first kinase inhibitor, imatinib (Gleevec), more than 44 kinase inhibitors were approved between 2001 and 2018 by the FDA [41]. Kinases are currently intensely studied due to the success of kinase inhibitors as anticancer agents; however, their lack of selectivity due to the similarities of different kinase binding sites remains a challenge. In order to facilitate the development of new kinase inhibitors, known binding sites on kinases were compared [42]. A total of 2,383 binding sites from 208 different kinase structures in the PDB were compared and the structural discriminants of the binding specificity and promiscuity of kinase ligands were

sought. Eight different inhibitor-binding regions in kinase binding sites were identified. The first four of these comprise the so-called "selectivity pocket", which depends on the type of kinase, while the remaining four are in the ATP binding site and contain highly conserved amino acids around the adenosine part of the ATP ligand. Although less specific, the ATP binding site nevertheless contains the so-called "selectivity entrance", which is different in different kinases, and allows the selectivity to be achieved also for the ATP-competitive inhibitors. It was concluded that, to achieve selectivity, new kinase inhibitors must be designed to bind to several regions in the selectivity pocket, while the ATP binding site excluding the selectivity entrance increases the promiscuity of these inhibitors.

4.2. Key proteins in disease-related networks

Polypharmacological strategies have emerged as the basis of a new approach to the modulation of disease networks. A pairwise comparison of 90,000 predicted binding pockets in 3,700 proteins was performed [40], and it was found that the space for polypharmacological opportunities is very large. As many as 23,148 protein pairs shared at least one similar binding site that could bind similar ligands. However, to achieve net effects on disease networks, multi-target drugs must modulate proteins with multiple functions. It was found that a protein shares a similar binding site with an average of seven proteins, among which are also those with unrelated structures and functions, and are therefore suitable as targets of multi-target drugs.

To achieve polypharmacological effects, it is necessary to focus on those sub-networks of the interactome that are related to a disease, and prioritize those proteins as targets, whose inhibition or removal highly affects the network stability. Therefore, key proteins with similar binding sites and different functions, were sought in disease-related networks. The analysis of the thrombotic network for example, revealed that simultaneous inhibition of known targets FGFR2 and KLK6 would result in a two-fold increase of the inhibitory effect compared to inhibition of a single target [40]. Furthermore, analysis of networks of cancer proteins revealed that multi-target drugs could

achieve on average a 2.7-fold increase in effectiveness. In stomach adenocarcinoma, skin cutaneous melanoma, and head and neck squamous cell carcinoma, the advantage of such multi-targeted drugs could be up to five-fold, and in ovarian cancer, simultaneous inhibition of CDK2, PPARG and ATRX proteins would result in a 2.1-fold improvement in comparison to the inhibition of ATRX alone. The study suggested that multi-target drugs targeting key proteins in protein-protein interaction networks with similar binding sites could achieve manifold improvements of the therapeutic effects when compared to single-target drugs.

4.3. Multi-chain proteins

Using the binding site comparison approach, it was found that multi-chain binding sites, those composed of several homomer protein chains, develop new functions more slowly than binding sites composed of a single chain only [43]. It was found that antibiotics and antiviral drugs targeting multi-chain binding sites enjoy a broader spectrum than drugs targeting single-chain binding sites. It was postulated that microorganisms develop resistance with more difficulty against drugs targeting multi-chain binding sites. Drug resistance was found to be particularly hindered when it was due to the accumulation of mutations, as is the case with rapidly evolving retroviruses. Consistent with the finding that multi-target drugs are more effective than single-target drugs in polygenic diseases such as cancers and psychiatric disorders [44,45], it was suggested that multi-chain binding sites are good target candidates for such multi-target therapies [46]. The retroviral protease inhibitor nelfinavir is a well-known example of a drug that binds in a multi-chain binding site of the HIV protease. It was suggested that due to this fact, nelfinavir is a potent polypharmacological agent and it has in fact been repurposed for treatment of cancer [47]. Multi-chain binding sites on homomeric proteins therefore appear to be promising as targets for multi-target drugs.

5. Emerging applications of binding site comparison in drug discovery

Some new emerging applications, described below of the binding site comparison approach have the potential to significantly impact drug development in the future.

5.1. Mapping of sequence variants to protein binding sites

Sequence variants are important due to their roles in various diseases as well as in the individual response to drugs. Single nucleotide polymorphisms (SNP), defined as replacements of a single nucleotide in genomes in >1% of the population are examples of sequence variants and SNPs that occur in protein coding regions result in altered proteins. Sequence variants are also somatic mutations, which influence cancer development [48]. Since binding sites determine protein function [49], the following questions were formulated: "Which sequence variants are inside binding sites?" and "Which sequence variants could disrupt ligand binding?" To answer these questions, homology docking and sequence-to-structure mapping was implemented in the GenProBiS web server [16], and provided an insight into the links between genes, binding sites, drugs, and diseases.

5.1.1 Determining effects of sequence variants on inhibitor binding

The enzyme indoleamine 2,3-dioxygenase catabolizes tryptophan and has an immunosuppressive role in cancer, which makes it an interesting target for the development of anticancer drugs [50]. The GenProBiS web server [16] was used to detect sequence variants that interact with known inhibitors of this enzyme. Two SNPs, rs764150078 (F163S) and rs774225205 (R231C), and one somatic mutation, COSM187719 (R231C), identified in the binding site were found to interact with the homology docked inhibitor N-(1,3-benzodioxol-5-yl)-2-{[5-(4-methylphenyl) [1,3] thiazolo [2,3-c] [1,2,4] triazol-3-yl] sulfanyl}acetamide [51]. It was shown that the F163S mutation eliminates the π–π stacking interaction between Phe163 and the imidazothiazole ring of the inhibitor, and that the R231C mutation vitiates the favorable electrostatic interaction between an arginine and the inhibitor at the entrance to the binding cavity [16]. The sequence variants that were

identified therefore probably lead to reduced binding of this inhibitor, and could, if the inhibitor entered clinical use, reduce its activity in patients containing these sequence variants in their genomes. It was suggested that routine sequencing of the patients' genomes could predict such cases and this approach thus represents a step towards precision medicine [16].

5.1.2 Explaining mechanisms of sequence variants on disease development

A mechanistic explanation for the observed links between sequence variants and diseases is often absent [52]. Gene TP53 for example, encodes the tumor suppressor protein p53 that plays a key role in the prevention of cancer as it maintains the stability of the genome. The TP53 is the most mutated gene (>50%) in human cancers, indicating that a potential cancer cell must bypass its regulation for the cancer to fully develop. The sequence variant rs121913343 (R273S) in the TP53 gene was associated with higher invasiveness of glioblastoma multiforme, an aggressive subtype of brain tumor, but its mode of action was unknown [53].

The variant was found to be in the protein-nucleic acid binding site on p53 identified using the GenProBiS web server [16] that predicts binding sites on proteins based on binding site comparisons. This binding site interacts with the response element located at the promoter of the apoptosis-related bcl-2-associated X protein gene. It was found that the replacement of Arg273 with serine eliminates the salt bridge between the p53 arginine and a phosphate group of the bound DNA, thus weakening the interaction between the p53 and the DNA. This sequence variant therefore acts negatively on apoptosis, since the bcl-2-associated X protein is a potent apoptosis activator [54]. The result is reduction of the activity of tumor suppression of the p53 protein. The same sequence variant is also located in an overlapping protein-protein binding site for the tumor suppressor p53-binding protein 2 (p53BP2) that in complex with p53, increases the binding of p53 to DNA and its transactivation function on pro-apoptotic gene promoters [55]. The mutation of Arg273 to serine was found to destroy a salt bridge interaction between the p53 arginine and the

glutamic acid (Glu1094) on the p53BP2, resulting, again, in reduced induction of apoptosis. The approach thus led to mechanistic insights into glioblastoma development.

5.2. Identifying targets of natural products

Natural products are the source of more than 32% of drugs in use today [56–58] and secondary metabolites of these natural products have a wide range of functions. Plants use them as weapons against predators or as attractants of symbiotic organisms. There is a current need to identify targets of secondary metabolites to explain their biological activities. Flavonoids for example, have diverse biological and health-promoting effects in cancer, inflammation and cytoprotectivity, which are achieved by influencing the activity of different signaling pathways.

Binding site comparison has been used successfully as a stand-alone method [59,60] or in combination with inverse virtual screening [61] to identify new flavonoid targets. Following the hypothesis that flavonoids leave biological imprints in the active sites of the biosynthetic enzymes, in which they are synthesized, a method was proposed to detect these imprints in the binding sites of potential target proteins [59]. Active sites of five representative flavonoid biosynthetic enzymes were compared with 8,077 druggable binding sites from the PDB. The flavonoid biosynthetic enzymes were chalcone isomerase (CHI) and chalcone synthase (CHS), from a flowering plant *Medicago sativa*, quercetin 2,3-dioxygenase (2,3QD) from the *Aspergillus japonicus* fungus, and dihydroflavonol-4-reductase (DFR) and leucoanthocyanidin reductase 1 (LAR) from the *Vitis vinifera canola* grape. These enzymes act on nine different substrates in five different metabolic pathways of flavonoid metabolism, and thus were assumed to represent the possible modes of flavonoid recognition. The screening achieved a significant enrichment of already known flavonoid targets with the Area Under the ROC Curve ranging between 0.68 and 0.78 depending on the biosynthetic enzyme that was used as the query. The flavonoid biological imprint, which was incorporated in the CHI enzyme, produced the most relevant hits. The CHI list contained many known flavonoid targets, including human RAC-α serine/threonine protein kinase [62], human

mitogen-activated protein kinase 1 [63] and human phosphatidylinositol 4,5-biphosphate 3-kinase catalytic subunit γ isoform [64], as well as many serine/threonine protein kinases. The diversity of flavonoid targets obtained using various flavonoid biosynthetic enzymes as queries indicated that the biological imprint obtained during the flavonoid biosynthesis is unique to each biosynthetic enzyme. It was suggested that it will be possible in the future to detect unknown targets of natural products that could then be biologically evaluated.

5.3. Identifying conserved structured waters in kinases

Water molecules bound to proteins often are involved in key interactions with drugs. Bosutinib, a type I kinase inhibitor in clinical use, forms such interactions with water molecules in the active site. These interactions significantly determine its selectivity towards certain kinases [65]. Consideration of water structurally conserved in active sites is therefore crucial for the development of new kinase inhibitors [66].

Binding site comparison was used to predict conserved structured water molecules on proto-oncogenic tyrosine-protein kinase Src bound to bosutinib [67]. In agreement with previous findings [65], two conserved water molecules W1 and W2 were discovered near the Met314 residue, which serves as a gatekeeper. Water W1 forms a hydrogen bond with the nitrile group of bosutinib, whereas water W2 does not interact directly with the inhibitor. Three new, previously unknown conserved structured water molecules were found, located near the bosutinib binding site. Water W3 binds to the indole group of Trp446 and the hydroxyl group of Tyr463 and could play an important role in conformational changes that occur when the γ-phosphate group of ATP is transferred to the tyrosine residue of the kinase substrate. The newly found waters, W4 and W5 are near the catalytically important Asp386 of Src kinase. These waters are bound between two β turns formed by the amino acid residues 384-387 and 405-408. Conserved waters are known to play a stabilizing role in twisted β turns [68,69]. It was suggested that the conserved waters at these sites are fully expected and could participate in the switching between active and inactive Src kinase

conformations [70]. Binding site comparison therefore provided an important insight into the structured and conserved water molecules, which should be valuable for the development of selective kinase inhibitors as well as for the study of kinase activity.

5.4. Modeling of the induced-fit effect to enable structure-based drug discovery

The MurA enzyme catalyzes the first step of cytoplasmic bacterial peptidoglycan synthesis and is an interesting target for the development of new antimicrobials [71]. However, the crystal structure of MurA from *Escherichia coli* in the PDB contains only apo binding sites, which proved challenging for structure-based drug discovery. It is recognized that molecular docking into holo binding sites gives higher enrichment of active compounds compared to docking into apo binding sites.

Homology docking of ligands and simulation of induced-fit effect was used to prepare the binding site on MurA enzyme structure [72]. In the process, this binding site became more holo-like, which permitted the discovery of new inhibitors of this enzyme. The binding site on the MurA from *E. coli* was superimposed with a similar binding site from the MurA from *E. cloacae*, whose holo form already existed in the PDB. The structure of the inhibitor which was co-crystallized with this holo protein was then transferred to the *E. coli* MurA. The comparison enabled superimposition of the precise apo and holo binding sites, and therefore revealed precisely the initial position of the transferred ligand in the apo binding site. Subsequently, the induced-fit effect was simulated so that the structure of the protein and the transferred ligand were energy minimized, eliminating steric clashes between the protein and the ligand. This generated a holo-like binding site on the *E. coli* MurA structure. Molecular docking to this enlarged binding site enabled the discovery of three new inhibitors of the MurA enzyme with novel scaffolds, the best with IC_{50} = 1 μM. Homology docking and induced-fit effect simulation thus enabled the preparation of a previously absent holo-like structure, which proved to be suitable for further structure-based drug development.

5.5. Modeling protein loops for biological drug design

Protein loops often have important roles in protein functions, protein-protein interactions, and protein-small molecule binding. Due to their flexibility, loops are often missing in the determined protein structures. Therefore, methods were developed [73–75] for loop homology modeling based on binding site comparisons. These methods search in the PDB for new loops that match the local query pattern around the missing protein loop. If such a loop is found, it is then used to modify the query protein, and is attached to the loop-flanking Cα atoms in the original structure. A query pattern can include the distance between the Cα atoms of the flanking residues or the structure and amino acid composition of the flanking segments.

Loop homology modeling is an important and evolving field. It was used to design biological drugs, such as therapeutic antibodies [76] and other therapeutically interesting proteins such as interleukin 34 [77]. It was also used to search for new antibody epitopes [78] obtained from polyclonal samples or engineered libraries, and to model missing loops in G-protein coupled receptors [79], which are targeted by approximately 34% of all modern drugs. Due to the success of therapeutic proteins, especially immunotherapeutics, we can expect even more diverse use of this approach in the future.

5.6. Finding common motifs in RNA tertiary structures to develop selective antibacterials

Like proteins, non-coding RNAs achieve their specific biological functions by folding into three-dimensional structures. RNA structural motifs have many important functions, for example, the kink-turn motifs on bacterial 23S ribosomal RNA (rRNA) are binding sites for nine proteins [80] and the universally conserved sarcin-ricin loop on 23S rRNA contains sites that are recognized and cleaved by ribotoxins [81]. Many methods for comparing RNA motifs exist [82–85] but comparison of RNA motifs is not as widely used in drug design as protein binding site comparison. Its use however may increase as more new disease roles of RNAs are discovered. In addition, there are

currently more than 7,000 protein-nucleic acid structures in the PDB, and this number is growing rapidly [33].

With the onset of resistance to ribosomal inhibitors, arguably the most successful antimicrobial agents, there is a renewed interest in the development of new protein synthesis inhibitors. Binding site comparison between the pathogenic and the non-pathogenic bacterial ribosomes revealed differences in the binding sites, which enabled the development of new erythromycin derivative antibiotics that specifically bind only to ribosomes of pathogenic species [86,87]. Similarly challenging is the lack of selectivity of tetracyclines. While these drugs act by inhibiting the 30S ribosomal subunit [88], it was recently discovered that they also influence translation by binding to the human 80S subunit [89]. Here, the comparison between human and bacterial rRNA could enable development of more specific tetracycline antibiotics or repurpose them as anticancer or anti-inflammatory drugs [89]. A methodological challenge is presented by thiopeptide antibiotics whose binding sites are composed of the rRNA as well as of the protein [90]. Currently there is no known method that would allow comparison of such composite binding sites.

6. Conclusion

It is important to recognize that most small molecule drugs bind to multiple proteins. A binding site comparison approach can predict and explain this multi-target drug binding. Such an approach involves a variety of computer methods that enable the detection of similarities between proteins independently of their sequences or folds. The pharmaceutical community has accepted this approach due to its applicability to drug repurposing and polypharmacology. However, new uses are emerging. In precision medicine, binding site comparison promises to explain the role of genetic variability on the development of diseases and effectiveness of drugs. In natural products, it enables finding unknown targets that could explain observed biological activities. It has enabled the discovery of conserved structured waters and simulations of induced-fit effects to facilitate structure-based drug discovery. In protein loop homology modeling, it has been applied to the

design of protein therapeutics and in RNA motif comparisons, an extension of this methodology to nucleoproteins could enable the design of more selective antibiotics.

7. Expert opinion

Binding site comparison plays an important role in drug development, in particular in the repurposing of drugs for new indications and in the related field of polypharmacology. The importance of the structural comparison of binding sites will only increase in the future due to the advancement of experimental methods for determination of protein structures, such as the electron microscopy, and the increasing accuracy of protein structure modeling [91].

The paradigm, “one drug, one target” is no longer valid. A drug can bind to several different proteins, which may have similar binding sites. On the one hand, this is an issue due to the side effects, and on the other, it is an opportunity for drug repurposing and to achieve synergistic effects with multi-target drugs. Drug repurposing will assist the discovery of drugs for rare and neglected diseases, and antibacterials, where *de novo* development is not economically viable. Polypharmacological drugs however, could significantly improve the treatment of complex diseases such as cancers and neurodegenerative disorders. Here, the binding site comparison approach is most powerful when it is used in combination with other virtual screening methods.

Binding site comparison enables proteome-wide prediction of binding sites and protein ligands. Together with the mapping of genes to binding sites, it provides mechanistic explanations of the biological effects of sequence variants. This could clarify the influence of sequence variants on disease development, and open the possibility of developing precision therapies.

Binding site comparison is showing promising results in the search for targets of drugs from natural products. The approach can be used independently as illustrated in the case of biological imprints, or in combination with other virtual screening methods. Natural products are the source of about 32% of all drugs in current use. They are abundant and their total synthesis is becoming increasingly feasible. With the use of innovative binding site comparison approaches, it is very

likely that this field will expand and perhaps produce some new drugs or new uses for existing drugs.

Binding site comparison also allows the identification of structured waters on proteins that play important roles in enzyme activity and drug discovery. Due to the precise local superimposition of protein structures, conserved water molecules outside the binding sites can now also be detected.

In combination with methods of molecular mechanics, binding site comparison allows the simulation of induced-fit effects, the adaptation of binding site structures caused by ligand binding, which enables structure-based drug discovery involving apo proteins.

Binding site comparison approach has been used to develop therapeutic proteins. It enables loop homology modeling, which is used in the development of biological drugs based on antibodies and other proteins. Due to the recent success of immunotherapeutics in cancer treatment, an ever-increasing and increasingly diverse use of this approach can be expected in the future.

Binding site comparison is currently limited to proteins, although many proteins also include nucleic acids as their integral parts. Some 5% of all structures in the PDB are complexes between proteins and nucleic acids that cannot be handled using current binding site comparison methods. The challenge is thus to extend the methodology to allow comparison of such composite binding sites. Such methods could, at the least, open up new opportunities for the development of new selective antibiotics.

In summary, comparison of binding sites is an established approach that grows as new experimental findings and techniques emerge. New applications presented in this review will be the basis for drug discovery and development in the future.

References

[1] Mirabello C, Wallner B. Topology independent structural matching discovers novel templates for protein interfaces. Bioinformatics. 2018;34:i787–i794.

[2] Konc J, Janežič D. ProBiS algorithm for detection of structurally similar protein binding sites by local structural alignment. Bioinformatics. 2010;26:1160–1168.

** An algorithm that detects novel binding sites on protein structures through local comparisons of whole protein structures without reference to existing binding sites.

[3] Xie L, Xie L, Bourne PE. A unified statistical model to support local sequence order independent similarity searching for ligand-binding sites and its application to genome-based drug discovery. Bioinformatics. 2009;25:i305–i312.

[4] Schalon C, Surgand J-S, Kellenberger E, et al. A simple and fuzzy method to align and compare druggable ligand-binding sites. Proteins: Structure, Function, and Bioinformatics. 2008;71:1755–1778.

[5] Kinoshita K, Furui J, Nakamura H. Identification of protein functions from a molecular surface database, eF-site. J Struct Func Genom. 2002;2:9–22.

[6] Wood DJ, Vlieg J de, Wagener M, et al. Pharmacophore Fingerprint-Based Approach to Binding Site Subpocket Similarity and Its Application to Bioisostere Replacement. J. Chem. Inf. Model. 2012;52:2031–2043.

[7] Schmitt S, Kuhn D, Klebe G. A New Method to Detect Related Function Among Proteins Independent of Sequence and Fold Homology. Journal of Molecular Biology. 2002;323:387–406.

[8] Brylinski M. eMatchSite: Sequence Order-Independent Structure Alignments of Ligand Binding Pockets in Protein Models. PLOS Computational Biology. 2014;10:e1003829.

[9] Yeturu K, Chandra N. PocketAlign A Novel Algorithm for Aligning Binding Sites in Protein Structures. J Chem Inf Model. 2011;51:1725–1736.

[10] Lee HS, Im W. G-LoSA for Prediction of Protein-Ligand Binding Sites and Structures. In: Kihara D, editor. Protein Function Prediction: Methods and Protocols [Internet]. New York, NY: Springer New York; 2017 [cited 2018 Dec 19]. p. 97–108. Available from: https://doi.org/10.1007/978-1-4939-7015-5_8.

[11] Aung Z, Tong JC. BSAlign: A rapid graph-based algorithm for detecting ligand binding sites in protein structures. Genome Informatics 2008 [Internet]. Imperial college press; 2008 [cited 2019 Feb 19]. p. 65–76. Available from: https://www.worldscientific.com/doi/abs/10.1142/9781848163324_0006.

[12] Chartier M, Najmanovich R. Detection of Binding Site Molecular Interaction Field Similarities. J Chem Inf Model. 2015;55:1600–1615.

[13] Rognan D. Chemogenomic approaches to rational drug design. Brit J Pharmacol. 2007;152:38–52.

[14] Baspinar A, Cukuroglu E, Nussinov R, et al. PRISM: a web server and repository for prediction of protein–protein interactions and modeling their 3D complexes. Nucleic Acids Res. 2014;42:W285–W289.

[15] Konc J, Janežič D. ProBiS-ligands: a web server for prediction of ligands by examination of protein binding sites. Nucleic Acids Res. 2014;42:W215–W220.

[16] Konc J, Skrlj B, Erzen N, et al. GenProBiS: web server for mapping of sequence variants to protein binding sites. Nucleic Acids Res. 2017;45:W253–W259.

** A web server that provides a link between genome sequences and protein structures and enables to determine which sequence variants are in protein binding sites and which could influence binding of ligands.

[17] Bhagavat R, Sankar S, Srinivasan N, et al. An Augmented Pocketome: Detection and Analysis of Small-Molecule Binding Pockets in Proteins of Known 3D Structure. Structure. 2018;26:499–512.

[18] Singh A, Kaushik R, Kuntal H, et al. PvaxDB: a comprehensive structural repository of Plasmodium vivax proteome. Database. 2018;2018:bay021.

[19] Naderi M, Govindaraj RG, Brylinski M. eModel-BDB: a database of comparative structure models of drug-target interactions from the Binding Database. Gigascience. 2018;7:giy091.

[20] Ehrt C, Brinkjost T, Koch O. Impact of binding site comparisons on medicinal chemistry and rational molecular design. J Med Chem. 2016;59:4121–4151.

** A dataset of test cases for binding site comparison approaches.

[21] Weber A, Casini A, Heine A, et al. Unexpected Nanomolar Inhibition of Carbonic Anhydrase by COX-2-Selective Celecoxib: New Pharmacological Opportunities Due to Related Binding Site Recognition. J. Med. Chem. 2004;47:550–557.

[22] Xie L, Evangelidis T, Xie L, et al. Drug Discovery Using Chemical Systems Biology: Weak Inhibition of Multiple Kinases May Contribute to the Anti-Cancer Effect of Nelfinavir. PLOS Computational Biology. 2011;7:e1002037.

[23] Naderi M, Lemoine JM, Govindaraj RG, et al. Binding site matching in rational drug design: algorithms and applications. Briefings in bioinformatics. 2018;bby078.

[24] Leinweber M, Fober T, Freisleben B. GPU-based point cloud superpositioning for structural comparisons of protein binding sites. IEEE/ACM transactions on computational biology and bioinformatics. 2018;15:740–752.

[25] Caffrey DR, Somaroo S, Hughes JD, et al. Are protein–protein interfaces more conserved in sequence than the rest of the protein surface? Protein Science. 2004;13:190–202.

[26] Konc J, Miller BT, Štular T, et al. ProBiS-CHARMMing: Web Interface for Prediction and Optimization of Ligands in Protein Binding Sites. J Chem Inf Model. 2015;55:2308–2314.

[27] Mysinger MM, Carchia M, Irwin JJ, et al. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. J Med Chem. 2012;55:6582–6594.

[28] Ehrt C, Brinkjost T, Koch O. A benchmark driven guide to binding site comparison: An exhaustive evaluation using tailor-made data sets (ProSPECCTs). PLOS Computational Biology. 2018;14:e1006483.

[29] Valencic E, Smid A, Jakopin Z, et al. Repositioning Drugs for Rare Immune Diseases: Hopes and Challenges for a Precision Medicine. Curr Med Chem. 2018;25:2764–2782.

[30] Chopra G, Samudrala R. Exploring Polypharmacology in Drug Discovery and Repurposing Using the CANDO Platform. Curr Pharm Design. 2016;22:3109–3123.

[31] Zhao Z, Martin C, Fan R, et al. Drug repurposing to target Ebola virus replication and virulence using structural systems pharmacology. BMC bioinformatics. 2016;17:90.

[32] Law V, Knox C, Djoumbou Y, et al. DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res. 2014;42:D1091–D1097.

[33] Burley SK, Berman HM, Christie C, et al. RCSB Protein Data Bank: Sustaining a living digital data resource that enables breakthroughs in scientific research and biomedical education. Protein Science. 2018;27:316–330.

[34] Rath A, Olry A, Dhombres F, et al. Representation of rare diseases in health information systems: The orphanet approach to serve a wide range of end users. Human Mutation. 2012;33:803–808.

[35] Joachim Haupt V, E Aguilar Uvalle J, Salentin S, et al. Computational Drug Repositioning by Target Hopping: A Use Case in Chagas Disease. Curr Pharm Design. 2016;22:3124–3134.

[36] Haupt VJ, Daminelli S, Schroeder M. Drug Promiscuity in PDB: Protein Binding Site Similarity Is Key. PLOS ONE. 2013;8:e65894.

[37] Brown N. Bioisosteres and Scaffold Hopping in Medicinal Chemistry. Mol Inform. 2014;33:458–462.

[38] Štular T, Lešnik S, Rozman K, et al. Discovery of Mycobacterium tuberculosis InhA Inhibitors by Binding Sites Comparison and Ligands Prediction. J Med Chem. 2016;59:11069–11078.

* A successful repurposing of existing antibacterials against *Mycobacterium tuberculosis* using binding site comparisons.

[39] Rawat R, Whitty A, Tonge PJ. The isoniazid-NAD adduct is a slow, tight-binding inhibitor of InhA, the Mycobacterium tuberculosis enoyl reductase: Adduct affinity and drug resistance. Proc Natl Acad Sci U S A. 2003;100:13881–13886.

[40] Duran-Frigola M, Siragusa L, Ruppin E, et al. Detecting similar binding pockets to enable systems polypharmacology. PLOS Comput Biol. 2017;13:e1005522.

[41] FDA-approved protein kinase inhibitors/US Food and Drug Administration approved small molecule protein kinase inhibitors [Internet]. [cited 2018 Dec 17]. Available from: http://www.brimr.org/PKI/PKIs.htm.

[42] Zhao Z, Xie L, Xie L, et al. Delineation of Polypharmacology across the Human Structural Kinome Using a Functional Site Interaction Fingerprint Approach. J Med Chem. 2016;59:4326–4341.

[43] Abrusán G, Marsh JA. Ligand Binding Site Structure Influences the Evolution of Protein Complex Function and Topology. Cell Reports. 2018;22:3265–3276.

[44] Peters J-U. Polypharmacology – Foe or Friend? J Med Chem. 2013;56:8955–8971.

[45] Roth BL, Sheffler DJ, Kroeze WK. Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. Nat Rev Drug Discov. 2004;3:353–359.

[46] Cairns RA, Harris IS, Mak TW. Regulation of cancer cell metabolism. Nat Rev Cancer. 2011;11:85–95.

[47] Koltai T. Nelfinavir and other protease inhibitors in cancer: mechanisms involved in anticancer activity. F1000Res [Internet]. 2015 [cited 2018 Dec 16];4. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4457118/.

[48] Forbes SA, Beare D, Boutselakis H, et al. COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res. 2017;45:D777–D783.

[49] Konc J, Janežič D. Binding site comparison for function prediction and pharmaceutical discovery. Curr Opin Struc Biol. 2014;25:34–39.

[50] Takamatsu M, Hirata A, Ohtaki H, et al. IDO1 Plays an Immunosuppressive Role in 2,4,6-Trinitrobenzene Sulfate–Induced Colitis in Mice. The Journal of Immunology. 2013;1203306.

[51] Tojo S, Kohno T, Tanaka T, et al. Crystal Structures and Structure–Activity Relationships of Imidazothiazole Derivatives as IDO1 Inhibitors. ACS Med Chem Lett. 2014;5:1119–1123.

[52] Miosge LA, Field MA, Sontani Y, et al. Comparison of predicted and actual consequences of missense mutations. Proc Natl Acad Sci U S A. 2015;112:E5189–E5198.

[53] Backes C, Harz C, Fischer U, et al. New insights into the genetics of glioblastoma multiforme by familial exome sequencing. Oncotarget. 2014;6:5918–5931.

[54] Oltval ZN, Milliman CL, Korsmeyer SJ. Bcl-2 heterodimerizes in vivo with a conserved homolog, Bax, that accelerates programed cell death. Cell. 1993;74:609–619.

[55] Samuels-Lev Y, O'Connor DJ, Bergamaschi D, et al. ASPP Proteins Specifically Stimulate the Apoptotic Function of p53. Molecular Cell. 2001;8:781–794.

[56] Baran PS. Natural Product Total Synthesis: As Exciting as Ever and Here To Stay. J Am Chem Soc. 2018;140:4751–4755.

[57] Ng SB, Kanagasundaram Y, Fan H, et al. The 160K Natural Organism Library, a unique resource for natural products research. Nat Biotechnol. 2018;36:570–573.

[58] Newman DJ, Cragg GM. Natural Products as Sources of New Drugs from 1981 to 2014. J Nat Prod. 2016;79:629–661.

[59] Sturm N, Quinn RJ, Kellenberger E. Similarity between Flavonoid Biosynthetic Enzymes and Flavonoid Protein Targets Captured by Three-Dimensional Computing Approach. Planta Med. 2015;81:467–473.

[60] Sturm N, Quinn RJ, Kellenberger E. Structural Searching of Biosynthetic Enzymes to Predict Protein Targets of Natural Products. Planta Med. 2018;84:304–310.

* Use of the binding site comparison methodology enabled the identification of targets for natural compounds.

[61] Carvalho D, Paulino M, Polticelli F, et al. Structural evidence of quercetin multi-target bioactivity: A reverse virtual screening strategy. Eur J Pharm Sci. 2017;106:393–403.

[62] El Amrani M, Lai D, Debbab A, et al. Protein Kinase and HDAC Inhibitors from the Endophytic Fungus Epicoccum nigrum. J Nat Prod. 2014;77:49–56.

[63] Tasdemir D, Mallon R, Greenstein M, et al. Aldisine Alkaloids from the Philippine Sponge Stylissa massa Are Potent Inhibitors of Mitogen-Activated Protein Kinase Kinase-1 (MEK-1). J Med Chem. 2002;45:529–532.

[64] Walker EH, Pacold ME, Perisic O, et al. Structural Determinants of Phosphoinositide 3-Kinase Inhibition by Wortmannin, LY294002, Quercetin, Myricetin, and Staurosporine. Molecular Cell. 2000;6:909–919.

[65] Levinson NM, Boxer SG. A conserved water-mediated hydrogen bond network defines bosutinib’s kinase selectivity. Nat Chem Biol. 2014;10:127–132.

[66] Klebe G. Applying thermodynamic profiling in lead finding and optimization. Nat Rev Drug Discov. 2015;14:95–110.

[67] Jukic M, Konc J, Gobec S, et al. Identification of Conserved Water Sites in Protein Structures for Drug Design. J Chem Inf Model. 2017;57:3094–3103.

[68] Rose GD, Young WB, Gierasch LM. Interior turns in globular proteins. Nature. 1983;304:654–657.

[69] Ogata K, Wodak SJ. Conserved water molecules in MHC class-I molecules and their putative structural and functional roles. Protein Eng Des Sel. 2002;15:697–705.

[70] Ozkirimli E, Post CB. Src kinase activation: A switched electrostatic network. Protein Sci. 2006;15:1051–1062.

[71] Jukic M, Gobec S, Sova M. Reaching toward underexplored targets in antibacterial drug design. Drug Development Research [Internet]. [cited 2018 Dec 16];0. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/ddr.21465.

[72] Rozman K, Lešnik S, Brus B, et al. Discovery of new MurA inhibitors using induced-fit simulation and docking. Bioorg Med Chem Lett. 2017;27:944–949.

[73] Karami Y, Guyon F, Vries SD, et al. DaReUS-Loop: accurate loop modeling using fragments from remote or unrelated proteins. Sci Rep. 2018;8:13673.

[74] Hildebrand PW, Goede A, Bauer RA, et al. SuperLooper—a prediction server for the modeling of loops in globular and membrane proteins. Nucleic Acids Res. 2009;37:W571–W574.

[75] Choi Y, Deane CM. FREAD revisited: Accurate loop structure prediction using a database search algorithm. Proteins: Structure, Function, and Bioinformatics. 2010;78:1431–1440.

[76] Kuroda D, Shirai H, Jacobson MP, et al. Computer-aided antibody design. Protein Eng Des Sel. 2012;25:507–522.

[77] Felix J, Elegheert J, Gutsche I, et al. Human IL-34 and CSF-1 Establish Structurally Similar Extracellular Assemblies with Their Common Hematopoietic Receptor. Structure. 2013;21:528–539.

[78] Hua CK, Gacerez AT, Sentman CL, et al. Computationally-driven identification of antibody epitopes. eLife [Internet]. 2017 [cited 2018 Dec 16];6. Available from: https://elifesciences.org/articles/29023.

[79] Nikolaev DM, Shtyrov AA, Panov MS, et al. A Comparative Study of Modern Homology Modeling Algorithms for Rhodopsin Structure Prediction. ACS Omega. 2018;3:7555–7566.

[80] Klein DJ, Schmeing TM, Moore PB, et al. The kink-turn: a new RNA secondary structure motif. The EMBO Journal. 2001;20:4214–4221.

[81] García-Ortega L, Álvarez-García E, Gavilanes JG, et al. Cleavage of the sarcin–ricin loop of 23S rRNA differentially affects EF-G and EF-Tu binding. Nucleic Acids Res. 2010;38:4108–4119.

[82] Duarte CM, Wadley LM, Pyle AM. RNA structure comparison, motif search and discovery using a reduced representation of RNA conformational space. Nucleic Acids Res. 2003;31:4755–4761.

[83] Ge P, Islam S, Zhong C, et al. De novo discovery of structural motifs in RNA 3D structures through clustering. Nucleic Acids Res. 2018;46:4783–4793.

[84] Piątkowski P, Jabłońska J, Żyła A, et al. SupeRNAlign: a new tool for flexible superposition of homologous RNA structures and inference of accurate structure-based sequence alignments. Nucleic Acids Res. 2017;45:e150–e150.

[85] Cruz JA, Westhof E. Sequence-based identification of 3D structural modules in RNA with RMDetect. Nat Methods. 2011;8:513–519.

[86] Eyal Z, Matzov D, Krupkin M, et al. Structural insights into species-specific features of the ribosome from the pathogen Staphylococcus aureus. Proc Natl Acad Sci U S A. 2015;112:E5805–E5814.

** Structural comparison of ribosome binding sites enabled the development of species-specific antibiotics.

[87] Matzov D, Bashan A, Yonath A. A Bright Future for Antibiotics? Annual Review of Biochemistry. 2017;86:567–583.

[88] Brodersen DE, Clemons WM, Carter AP, et al. The Structural Basis for the Action of the Antibiotics Tetracycline, Pactamycin, and Hygromycin B on the 30S Ribosomal Subunit. Cell. 2000;103:1143–1154.

[89] Mortison JD, Schenone M, Myers JA, et al. Tetracyclines Modify Translation by Targeting Key Human rRNA Substructures. Cell Chemical Biology [Internet]. 2018 [cited 2018 Dec 16]; Available from: http://www.sciencedirect.com/science/article/pii/S2451945618303301.

[90] Lin J, Zhou D, Steitz TA, et al. Ribosome-Targeting Antibiotics: Modes of Action, Mechanisms of Resistance, and Implications for Drug Design. Annual Review of Biochemistry. 2018;87:451–478.

[91] Richard Evans, John Jumper, James Kirkpatrick, et al. De novo structure prediction with deep-learning based scoring [Internet]. 2018. Available from: http://predictioncenter.org/casp13/doc/CASP13_Abstracts.pdf.