JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Large-Scale Mining for Similar Protein Binding Pockets: With RAPMAD Retrieval on the Fly Becomes Real

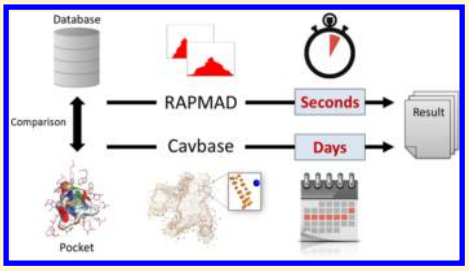
Timo Krotzky,[†] Christian Grunwald,[‡] Ute Egerland,[‡] and Gerhard Klebe*[,†]

[†]Department of Pharmaceutical Chemistry, Philipps-Universität Marburg, Marbacher Weg 6-10, 35032 Marburg, Germany

[‡]BioCrea GmbH, Meissner Strasse 191, 01445 Radebeul, Germany

Supporting Information

ABSTRACT: Determination of structural similarities between protein binding pockets is an important challenge in *in silico* drug design. It can help to understand selectivity considerations, predict unexpected ligand cross-reactivity, and support the putative annotation of function to orphan proteins. To this end, Cavbase was developed as a tool for the automated detection, storage, and classification of putative protein binding sites. In this context, binding sites are characterized as sets of pseudocenters, which denote surface-exposed physicochemical properties, and can be used to enable mutual binding site comparisons. However, these comparisons tend to be computationally very demanding and often lead to very slow computations of the similarity measures. In this study, we propose RAPMAD (RApid Pocket MAtching using Distances), a new evaluation formalism for Cavbase entries that allows for ultrafast similarity comparisons. Protein binding sites are represented by sets of distance histograms that are both generated and compared with linear complexity. Attaining a speed of more than 20 000 comparisons per second, screenings across large data sets and even entire databases become easily feasible. We demonstrate the discriminative power and the short runtime by performing several classification and retrieval experiments. RAPMAD attains better success rates than the comparison formalism originally implemented into Cavbase or several alternative approaches developed in recent time, while requiring only a fraction of their runtime. The pratical use of our method is finally proven by a successful prospective virtual screening study that aims for the identification of novel inhibitors of the NMDA receptor.



INTRODUCTION

Prediction of protein function and attempts to explain putative cross-reactivities of drug molecules are key objectives in biological sciences in general and pharmaceutical drug development in particular. The cross-binding of a given drug at an unexpected protein target (so-called off-target effects) are one of the major explanations for adverse drug reactions (ADRs). As outlined by Sim and Ingelman-Sundberg,[1] ADRs account for 7% of all hospitalizations, 20% of readmissions to hospitals, and 4% of withdrawals of new chemical substances. Hence, they are at least as costly as the drug treatment itself, and they can be rather ruinous at late-stage drug discovery. With 100 000 lethal cases per year, ADRs are among the leading causes of death in the United States.[1] In addition, other studies demonstrate that a single drug molecule interacts on average with six targets in a cell in addition to the desired one.[2] This underlines the high risk for unwanted cross-reactivities of a newly developed drug. In consequence, the prediction of ADRs by quantifying a drug's risk to interact with unexpected proteins is of utmost importance already at early stages of the drug discovery process, preferentially long before expensive clinical studies are conducted.

Finding similar pockets in the entire protein binding pocket space that can potentially host different ligands will provide valuable information during the lead optimization phase while improving drug binding in terms of higher affinity and selectivity. Following such a strategy, tailored design of promising candidate molecules featuring molecular decorations, incorporating bioisosteric replacements, or even pursuing the exchange for novel scaffolds (*scaffold-hopping*) can be acquired. Along completely different lines, pocket comparisons can also be exploited to successfully annotate biochemical function to orphan proteins.[3]

Identifying druggable binding sites is the initial task in structure-based drug discovery. The continuously growing number of available protein structures has increased the desire for automated detection and comparison algorithms to assign putative binding pockets. To accomplish this task, several binding site detection algorithms have emerged to unravel the so-called pocketome,[4−7] which can be basically classified in terms of geometry-based and energy-based methods. In addition, several binding site comparison protocols have been proposed, which are based on the comparative evaluation of structural information. In general, the methods reported in the literature to compare proteins fall into three categories: fold-based, template-based, and surface-based.[8] The fold-based methods include, for example, DALI[9] and MC-CE,[10] as well as SABERTOOTH[11] and CATHEDRAL.[12] Meanwhile, several fold databases have been developed based on these

Received: September 29, 2014
Published: December 4, 2014

dx.doi.org/10.1021/ci5005898 | *J. Chem. Inf. Model.* 2015, 55, 165−179

methods.[13−15] Template- and surface-based methods do not compare entire proteins but evaluate only parts of their structure. This strategy is based on the assumption that functionally important regions are evolutionary conserved and, thus, more relevant for the mutual comparison. Template-based methods comprise algorithms such as PINTS,[16] JESS,[17] or LabelHASH.[18]

Zauhar et al.[19] developed a surface-based method, which builds on a technique called *shape signatures* to describe either the shape of the bound ligand molecules and of the corresponding receptor sites. Here, the volume of the protein binding site is explored by a ray-tracing method. Probability distributions of surface-based descriptors can be derived by this procedure, and they are stored in terms of histograms. Subsequently, these histograms are used to test for a shape complementary between compounds and receptors. Binkowski and Joachimiak[20] developed a two-step procedure for the comparison of binding site surfaces that consists of global shape matching based on distance calculations between all unique atom pairs followed by a spatial alignment of the physicochemical texture to identify conserved amino acids. Furthermore, prominent surface-based approaches include methods such as CASTp,[21] EF-Site,[22] SiteEngine,[23] and Cavbase.[24,25] The latter group of tools is of special interest in the context of structure-based drug design. Here, the basic idea is to capture physicochemical properties of functional groups that are essential for the interaction of proteins and ligands. The goal of these approaches is to identify portions of proteins (binding sites) that are likely to recognize and thus interact with similar ligand molecules, independent of how much they actually vary in overall amino acid sequence.[26]

In this study, we focus on Cavbase, a module of the Relibase+ system[27,28] that was developed for storing information about putative protein binding sites. Cavbase exploits the grid-based LIGSITE algorithm[29] to detect cavities exposed to the surface of proteins stored in the Protein Data Bank (PDB).[30] To be considered in the database, a detected depression must comprise a cluster of more than 320 adjacent grid intersections (grid spacing of 0.5 Å resulting in about 40 Å^3), and it must exhibit a certain minimal penetration depth. Next, the putative binding sites are classified in terms of a sparse set of pseudocenters, leading to a compressed representation. The pseudocenters are assigned to atoms or groups of atoms of the amino acids that flank the detected cavity with a distance from the rim closer than 1.1 Å. They cipher physicochemical properties that are important for molecular recognition next to the adjacent surface patch, potentially forming an interaction to the bound ligand. In total, seven different pseudocenter types are considered: H-bond donors, H-bond acceptors, H-bond donors/acceptors (that we will call "doneptors"), centers of aromatic rings, centers comprising pi electrons, aliphatic groups, and metal ions (see Figure 1 for an example).

One important feature of Cavbase is the option to mutually compare binding sites. While such a similarity consideration of entire proteins is easily performed on the sequence level by applying alignment methods like BLAST,[32] a comparison on the structural level becomes computationally a much more demanding but required task. Proteins may exhibit similar functions, even though they are only sharing a low sequence similarity and overall fold.[33,34]

To measure the structural similarity between two protein binding sites, Cavbase follows, in the currect implementation, a

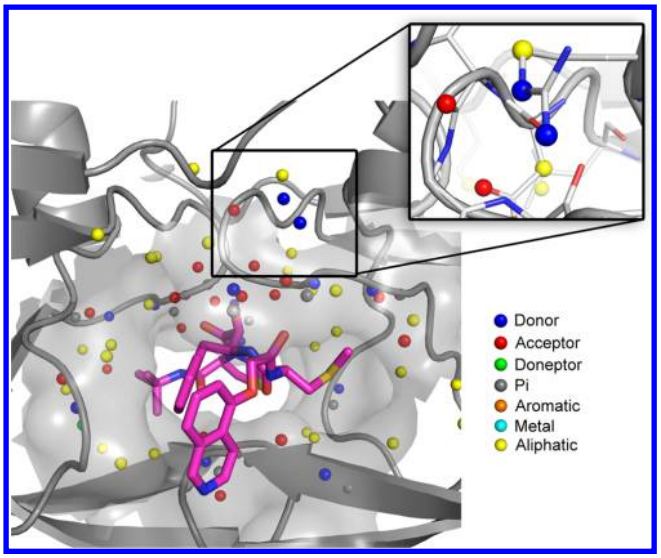


Figure 1. Example of binding site representation of HIV protease in Cavbase (PDB code 1HPX). A pocket is denoted by a set of pseudocenters in 3D space (colored spheres). Pseudocenters are assigned to specific parts of the amino acids flanking the cavity and represent their physicochemical properties, which is shown as a blow-up inset on the upper right side. The ligand (carbon atoms, magenta) is shown for clarity. The illustration was prepared using PyMOL.[31]

two-step procedure. At first, the two sets of pseudocenters are matched to find reasonable mutual superpositions. To this end, the pseudocenter sets are transformed into undirected graphs, which are then used to build a product graph. Afterward, the Bron-Kerbosch algorithm[35] is applied on this product graph to detect the 100 largest cliques, each representing a possible superposition of the two pseudocenter subsets. As the problem of clique detection is known to be NP-complete and therefore computationally quite demanding, Cavbase is limited to the detection of this subset. To compensate for this simplification, it performs a second computational step to evaluate shape similarity. Every generated superposition is evaluated by considering an expanded set of surface points, which describe the putative contact surface of the protein. This second step is again computationally rather time consuming. The final similarity score is calculated by assessing the maximal degree of overlapping substructures of two binding sites.

Although some alternative approaches for the comparison of Cavbase input binding site data have been developed,[36−38] the overall comparison remains a computationally intensive task. Many methods suffer from the NP-complete optimization problem, and even the most efficient ones exhibit a cubic time complexity.[37,38] This often results from the representation of protein binding sites by means of graphs, which usually require solutions by least-squares techniques. Even though recently published methods such as SEGA,[38] GAVEO,[36] or LPCS[37] exhibit impressive accuracy, their long runtime still excludes them from large-scale applications or interactive modeling considerations.

To reduce the runtime of such comparisons, we investigated an alternative representation of Cavbase entries. More efficient approaches employ distance histograms as a means to describe binding sites. Recently, Behren et al.[39] presented TrixP, an index-based method that is able to perform fast protein binding site comparisons. TrixP uses descriptors that encode pharmacophoric as well as spatial features to determine binding

site similarities by employing a partial shape matching. After this matching step, the binding sites are superimposed, and finally, similarity is assessed by means of the overlay of pharmacophoric properties. The binding sites can be provided by either defining a reference ligand or by employing the built-in DoGSite method[40] to predict putative binding sites. Yeturu and Chandra[41] proposed a method called PocketMatch that implements a binding site representation by 90 lists of sorted distances, which are later aligned to accomplish a comparison. The authors were able to show that this representation is appropriate to combine information regarding shape and physicochemical properties. Another example is BSSF, a fingerprint concept, proposed by Xiong et al.[42] The fingerprints are used to store frequencies of spatial distances between the centroids of key interaction portions of the pocket's amino acids. A total of 49 fingerprints, each containing 41 integer values, are generated and subsequently compared by using the Canberra distance measure. Weill and Rognan[43] have developed FuzCav, a method that is also using integer vectors to describe protein binding sites (4833 vectors). This fingerprint stores the frequencies of pharmacophoric triplets that were generated from the Cα atoms of amino acids flanking the binding pocket. Comparisons by means of FuzCav can be executed very efficiently due to the ultrafast comparison of fingerprint data structures. For evaluation, binding sites of the sc-PDB database[44] were exploited. Most important, in all three mentioned methods, a putative binding site is reduced by considering only the local environment around a given ligand mainly to avoid noise during the automated pocket detection. A distance of 4 Å (PocketMatch), 6 Å (BSSF), and 6.5 Å (FuzCav) is used for the extraction, and comparison procedures are then applied on these pockets, which are well-defined but clearly biased by an initially bound ligand. In contrast, we will present a histogram method that is instead based on the entirely automatic detection of binding pockets in Cavbase, unbiased by the prerequisite to show a bound ligand in the input structure.

We propose the description of binding pockets by a set of distance histograms, which can be rapidly generated and compared, since both tasks are conducted with linear complexity. This allows for an ultrafast binding site comparison, where geometric shape and physicochemical features of the cavities are considered simultaneously, as the combination of both descriptors is highly relevant to characterize binding sites.[45]

MATERIALS AND METHODS

Histogram Representations of Binding Sites. The workflow of the RAPMAD approach is shown in Figure 2 and will be explained in detail. First, the assigned set of pocket pseudocenters P is split into seven subsets $P_1 \ldots P_7$, according to the pseudocenter classification scheme based on the physicochemical properties (step 1). Next, two spatial reference points are defined for each subset, namely, the centroid and the centroid closest (step 2). Such points have already been used by Ballester and Richards[46] to perform efficient comparisons of small molecules. Osada et al.[47] also used the centroid to derive shape distributions of various 3D objects and facilitate fast subsequent comparisons. The centroid c represents the geometric center of all pseudocenters in a subset P_n, and its coordinates c_x, c_y, and c_z are defined as

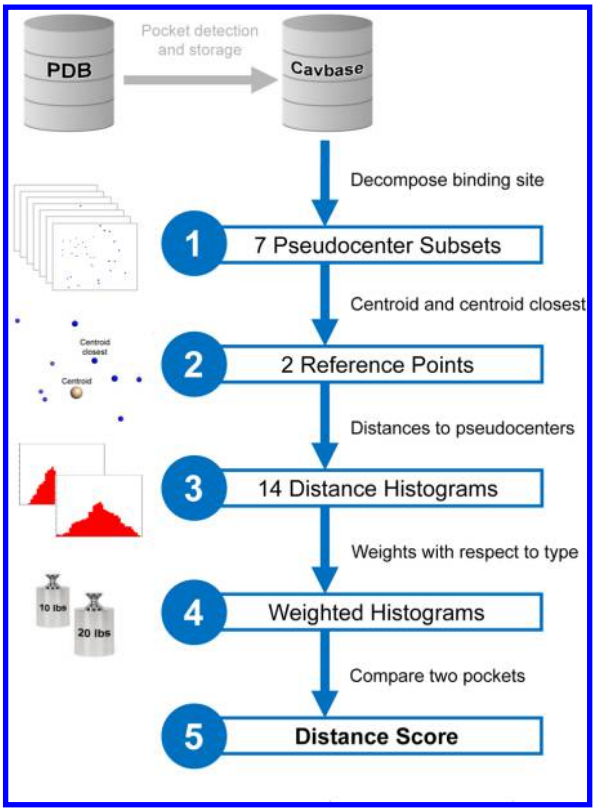


Figure 2. Workflow of the RAPMAD approach (highlighted in blue). In the first step, the set of pseudocenters that represents a binding site is divided into seven subsets based on their physicochemical properties. Subsequently, two reference points (centroid and centroid closest) are determined for every subset (step 2). In step 3, distances from these reference points to all neighboring pseudocenters are calculated and summarized in individual histograms. The histograms are subsequently weighted with respect to the relative pseudocenter frequencies in Cavbase (step 4). Finally, the histograms are used to accomplish a pairwise pocket comparison (step 5).

$$c_x = \frac{\sum_{p \in P_n} p_x}{|P_n|}$$

$$c_y = \frac{\sum_{p \in P_n} p_y}{|P_n|}$$

$$c_z = \frac{\sum_{p \in P_n} p_z}{|P_n|}$$

While c is an assigned point in space not necessarily coinciding with any atom or pseudocenter of the molecule, the centroid closest cc is one pseudocenter of the subset that shows the smallest Euclidean distance to c.

For every subset, the distances from these two reference points to all other pseudocenters $p \in P$ are calculated and summarized in terms of histograms using a bin size of 0.4 Å (step 3). This way, a spatial distribution profile of all pseudocenters in the pocket as well as the relative position of a subset within the total set of pseudocenters are captured. To find the best performing bin size, we examined all values in the interval [0.1,2.0] Å with a step size of 0.1 Å and found 0.4 Å to work most properly. Figure 3 shows RAPMAD's accuracies when using varying bin sizes in the interval [0.1,2.0] Å with a step size of 0.1. The values represent the success rates in a

dx.doi.org/10.1021/ci5005898 | J. Chem. Inf. Model. 2015, 55, 165−179

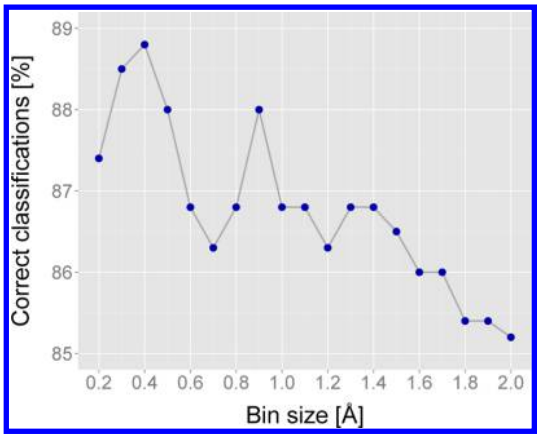


Figure 3. Variation of RAPMAD's accuracy when changing the histogram bin sizes in the interval [0.1,2.0] Å. The best classification result is obtained with a bin size of 0.4 Å. Increasing the bin size leads to a nearly continuous deterioration of the success rates. The worst value is, however, obtained when using a bin size of 0.1 Å (75.6%, not shown in the diagram).

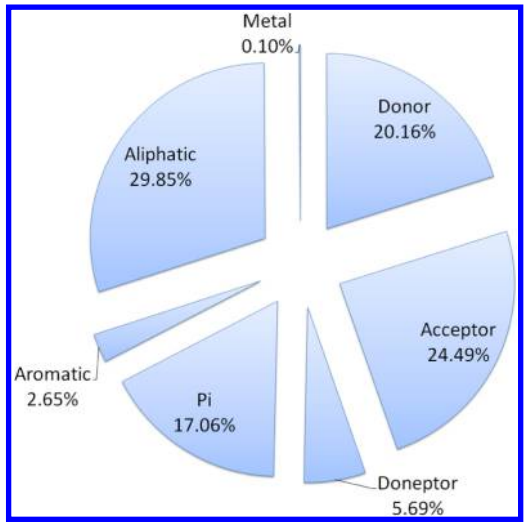


Figure 4. Frequency of the different pseudocenter types in Cavbase entries. We evaluated the entire database, consisting of 25 452 124 pseudocenters out of 275 097 putative binding sites. The vast majority of all pseudocenters are hydrogen-bond acceptors, hydrogen-bond donors, and aliphatic centers. Taken together, they account for almost 75%. Clearly less frequent are pseudocenters representing pi centers and H-bond doneptors. Also, the set of aromatic pseudocenters is rather sparsely populated (2.65%). The least frequent, however, is the type metal centers that only accounts for about 0.1%.

classification experiment on a test set of pockets that bind either ATP or NAD called ATP/NAD_{small} (see Data Sets section). Apart from *c* and *cc* as representative reference points, we also examined other definitions of reference points, for example, the centroid-furthest pseudocenter. As a result, the selection of *c* and *cc* turned out to be the most information-rich one with respect to efficiency.

In total, a set of 14 histograms (two histograms for each of the seven pseudocenter types) is generated to describe pockets. In addition, we assign a weight to each histogram to consider the relative occurrence frequency of the respective pseudocenter type in Cavbase (step 4). According to Xiong et al.,[42] who stated that donor, acceptor, doneptor, and aromatic functionalities are enriched in binding sites, we selected a weighting scheme that accounts for this observation. All 25 million pseudocenters in Cavbase were evaluated. As shown in Figure 4, the most frequent pseudocenters (29.9%) are assigned to an aliphatic functionality and will hence receive the least weight. Also donors and acceptors are quite commonly populated with 20.2% and 24.5%, respectively. The smallest fraction is the set of metal pseudocenters, which only accounts for 0.1%.

Weights are determined in a way that is often referred to as the inverted Boltzmann method.[48] It is supposed to weight those pseudocenters highest that occur least in the given data sample, as it is assumed that they are carrying most of the discriminating information. The weight *w* for every histogram is thus calculated as

$$w = \log\left(\frac{1}{f}\right)$$

where *f* denotes the relative frequency of a certain pseudocenter type.

We also calculated the average histograms for each pseudocenter type in Cavbase (14 histograms in sum) to find the areas with highest variation. Bins in the range from 20 to 50 that represent distances from 8 to 20 Å exhibited the highest standard deviations, which were thus incorporated to accomplish another bin-wise weighting. However, it was not possible to further improve the results by this procedure, we therefore decided to remain with the above-mentioned histogram weighting.

Pocket Comparison. The comparison of two pockets *A* and *B* (step 5) is performed by mutually facing two sets of the 14 histograms. Each pair of histograms assigned to the same pseudocenter type is matched, which leads in total to 14 pairwise comparisons. We examined a large variety of different histogram comparison methods such as the Minkowski distance,[49] Kolmogorov–Smirnov test,[50,51] Jaccard measure,[52] Hellinger distance,[53] or Canberra distance.[54] Finally, we decided to use the Jensen–Shannon divergence[55] to carry out the required comparisons, as this method provided the highest accuracy and can be calculated efficiently. To facilitate the pairwise comparison of two histograms, the smaller one is always extended with empty bins. For two histograms H_A and H_B the Jensen–Shannon divergence is based on the Kullback–Leibler divergences[56] of (H_A,M) and (H_B,M), where *M* is the median distribution of H_A and H_B. The value of a bin *i* in *M* is calculated as follows.

$$M[i] = \frac{H_A[i] + H_B[i]}{2}$$

The Kullback–Leibler divergence KLD of this median distribution *M* and a histogram *H* is then defined as

$$\mathrm{KLD} = \sum_i \log\left(\frac{H[i]}{M[i]}\right) \times H[i] \times H_w$$

where H_w returns the weight of *H*. If the value of any bin *i* of either *H* or *M* is zero, it is discarded. Having calculated KLD_A for (H_A,M) as well as KLD_B for (H_B,M), the final Jensen–Shannon divergence JSD can be computed as

$$\mathrm{JSD} = \frac{\mathrm{KLD}_A + \mathrm{KLD}_B}{2}$$

dx.doi.org/10.1021/ci5005898 | J. Chem. Inf. Model. 2015, 55, 165–179

After the JSD was calculated for every pairwise histogram comparison, all 14 values are summed up to determine the final distance score S of the two binding sites.

$$S = \sum_{i=1}^{14} \mathrm{JSD}(H_{Ai}, H_{Bi})$$

Obviously, the lowest distance scores are obtained if a pocket is compared with itself, which leads to a value of zero. In general, S falls in the interval $[0,\infty)$ because a maximal distance of two histograms in terms of the JSD cannot be defined.

We also investigated the influence of normalizing and smoothing the histograms previous to the comparisons. In a normalization, the integral of every histographic distribution is calibrated to one. It has been frequently applied to reduce the impact of the numbers of elements in the distribution (numbers of pseudocenters in a pocket) on shape comparisons.[45] Smoothing functions, for example, Gaussian smoothing,[57] triangular smoothing,[58] or a sliding window averaging,[59] are often applied to counteract the discontinuous behavior of conventional histograms[60] as well as crystallographic uncertainties. It turned out, however, that both methods were not able to improve our results, which was why we rather omitted these additional steps and achieved another runtime gain instead.

Data Sets. *Pockets Binding a Particular Ligand.* To evaluate our novel comparison procedure, we used several validation sets of protein binding sites, which will be described in the following and which have been considered by others for the same purpose. As reported by Fober et al.,[36] the first applied data set was originally assembled to examine the discriminative power of the graph-based comparison method GAVEO. A set of 355 binding pockets hosting either the cofactor nicotinamide-adenine-dinucleotide or adenosine-5′-triphosphate (PDB ligand identifiers NAD and ATP) was collected. Because a protein may exhibit several binding sites for the ligand, the number of selected pockets per protein was restricted, so that each protein was considered only once in the data set. Furthermore, the number of binding sites was reduced by calculating the root-mean-square deviation (RMSD) between ligand pairs occupying the binding sites. By defining a maximal RMSD threshold of 0.4 Å, this step filters for pockets only hosting cofactors adopting similar conformations. Finally, these selection criteria resulted in a set of 141 ATP- and 214 NAD-binding pockets. This data set will be referred to as ATP/$\mathrm{NAD}_{\mathrm{small}}$.

In addition, we compiled another larger and more comprehensive data set of ATP- and NAD-binding pockets now omitting the above-mentioned RMSD constraint. In consequence, this set of pockets also hosts ligands with diverse conformations. Moreover, we retrieved a set of binding sites from Cavbase that accommodate flavin-adenine-dinucleotide (PDB ligand identifier FAD), another cofactor related to NAD and ATP that is used by many enzymes in biology. Sets of pockets hosting particular ligands such as ATP, NAD, or FAD were chosen to establish challenging comparisons, as Stegemann and Klebe[61] showed that these cofactors are able to bind in various orientations, even though sharing the adenosine diphosphate moiety as common substructure. As sole constraint, we decided not to consider binding pockets with a volume greater than 4000 $\mathrm{Å}^3$ in the data set. Regarding a value of 1.7 Å as van der Waals radius of a carbon atom, such pockets would still provide enough space to host approximately 150 atoms. In consequence, we assume that any larger cavity will certainly display an artificially extracted pocket falsely selected by the LIGSITE algorithm. For the data set ATP/$\mathrm{NAD}_{\mathrm{large}}$, we compiled a sample of 420 ATP- and 402 NAD-binding pockets, now also showing a better balanced ratio between the two class sizes. This provides an additional challenge to our classification experiments, as the success rate of a simple randomized assignment tends to improve in case of an unbalanced data set with nonequal class sizes.[62] Classification rates that result from a majority voting achieve already 214/355 = 60.28% for a simple random assignment using the ATP/$\mathrm{NAD}_{\mathrm{small}}$ data set. In the new data set, this voting reduces to 420/822 = 51.09%.

The FAD data set was supposed to be even more challenging because for this cofactor it has been stated that no single protein-based pharmacophore can be derived using binding pocket information.[63] Hence, we expected the set of FAD-binding pockets to be rather diverse with respect to arrangements of residues interacting with the ligand. In this set, we furthermore distinguished between a covalently bound and a noncovalently bound FAD. In total, we selected 429 pockets binding FAD noncovalently and 114 pockets hosting the cofactor via covalent attachment. We refer to these data sets as FAD and $\mathrm{FAD}_{\mathrm{cov}}$, respectively.

Data Sets Based on EC Classes. We also considered two data sets comprising protein binding sites classified by the EC class numbers, the corresponding enzymes to which they have been assigned. First, we used a diverse data set of 502 pockets compiled by Glinca and Klebe, which spreads over all six main EC classes.[64] Table 1 provides an overview of this compilation, which we name EC Data Set.

Table 1. EC Data Set Comprises 502 Protein Binding Sites That Were Arranged in 16 Subsets, Which Cover All Six Main EC Classes

EC class	name	number of pockets
1.1.1.21	aldose/xylose reductase	62
1.1.1.42	isocitrate dehydrogenase	21
1.1.1.62	estradiol 17β-dehydrogenase	16
1.14.13.2	hydroxybenzoate-monooxygenase	30
2.7.1.37	cyclin-dependent kinase 2	46
2.7.1.112	C-Src tyrosine kinase	20
2.7.4.9	thymidylate kinase	35
3.4.21.5	thrombin	41
3.4.23.16	HIV-1 protease	48
3.4.24.86	TNF-α converting enzyme	16
4.1.1.23	COMP-decarboxylase	36
4.2.1.1	α-carbonic anhydrase I, II, III, IV	70
5.3.1.5	xylose isomerase	13
5.4.2.1	phosphoglycerate mutase	5
6.3.2.1	pantoate-β-alanine ligase	27
6.3.4.4	adenylosuccinate synthase	16

In addition, we used an EC class data set that consists of 1028 binding sites of serine proteases. The set was compiled in-house by two of our collaborators (Strickert and Fober, personal communication, see Acknowledgments). As pointed out by Schomburg and Rarey,[65] serine proteases possess a highly conserved overall sequence and structural similarity. In particular, the catalytic triad of serine, histidine, and aspartate that represents the key sequence motif in the active site is present in all considered structures. As stated by the authors,

dx.doi.org/10.1021/ci5005898 | J. Chem. Inf. Model. 2015, 55, 165−179

any sequence-based classification methods will therefore most likely fail on this target class. Furthermore, Schomburg and Rarey underscore the importance of the serine proteases in drug development by stressing that the selective inhibition of thrombin and factor Xa is of high pharmacological interest with respect to a direct anticoagulant therapy.

The binding sites in this *serine protease data set* are distributed over 40 subclasses of the EC class 3.4.21. This set also includes a subclass "0", to which all pockets were artificially assigned that were so far not attributed to any of the other classes (Figure 5). As for all data sets used in the present study, each PDB entry was restricted to one single pocket and therefore it appears only once in the set.

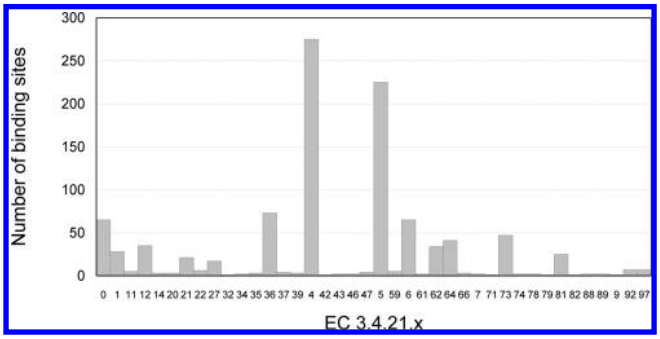


Figure 5. Distribution of binding sites in the serine protease data set. A total of 1028 structures are arranged in 40 subclasses of EC class 3.4.21, which are listed along the *x*-axis. Subclass 0 comprises all the binding sites (65 entries) that could not be annotated to any particular subclass of 3.4.21.x. The highest populated subclasses 4 and 5 contain pockets of trypsin and thrombin structures, respectively.

Proteases. Finally, we used another data set previously studied by Glinca and Klebe,[64] which comprises 90 binding sites of proteases and is termed *protease data set*. Also, for this data set, binding sites were selected that comprise the catalytic centers of enzymes. The authors used this data set as a challenge for a clustering on the Merops clan level, which worked well using the similarity measure of Cavbase but completely fails when using approaches solely based on sequence comparisons. We can therefore rule out any remarkable redundancies with respect to sequence space similarity in this data set. By using the latter set of binding sites, we will be particularly able to assess our comparison procedure with regard to conserved three-dimensional substructures. Merops[66] clans have been predominantly assigned to capture such information via the recognition of shared substrate motifs. Thus, a comparison with respect to the Merops classification will enable further assessment of our novel comparative approach. More precisely, we will use the Merops clans as templates for another classification test. Subsequently, this will allow us to check to what extent our results agree to the independently assigned Merops classification.

Evaluation Techniques. *ROC Curves.* A commonly used technique in signal detection and medical statistics to test whether a new method is able to distinguish between hits and decoys particularly with respect to retrieval rates are receiver operating characteristic (ROC) curves.[67,68] ROC curves depict the relative trade-off between success and failure rates by plotting the number of true positives (TP) against the number of false positives (FP). Correct entries that are falsely recognized as decoys are named "false negatives" (FN), and those that are correctly retrieved as decoys are classified as "true negatives" (TN). A ROC curve plots the true positive rate (TPR, also called recall rate) on the *y*-axis against the false positive rate (FPR) on the *x*-axis, where the TPR and FPR are defined as

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

and

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{TN} + \mathrm{FP}}$$

Each ROC curve starts at the origin (0,0), and a perfect search would result in full retrieval on first ranks (0,1). In this case, the graph would possess an infinitely high gradient, resulting in area under the curve (AUC) of 1. The AUC is a prominent descriptor to rank the quality of a method. A retrieval with a random selection of hits will lead to a ROC curve showing unit slope and an AUC of 0.5.

In this contribution, we will use ROC curves to evaluate the results of our retrieval experiments. For instance, several pockets that bind a particular ligand are compared against the entire Cavbase database (version 2.2.2, July 2010) comprising 275 097 pockets. All pockets binding the same ligand will be defined as a hit and all others as decoys. In our experiments, we used cavities binding the chemotherapeutic agents methotrexate and pemetrexed as queries. It is generally assumed that a method performing reasonably well in such a retrospective virtual screening is also likely to succeed in a prospective screening scenario.[48]

Cross-Validations. While comparing molecular structures such as binding sites, it appears rather difficult to assess calculated values such as similarity scores directly, as these scores do not return a kind of threshold value defining whether two pockets are "similar" or "dissimilar". To overcome this problem, we make use of an indirect measure by retrieving entries that are closest to the query in terms of distance in score space. Next, a property of interest is extracted from the list of *nearest neighbors* (NN). In the present study as property the name of the bound ligand was compared to that used in the query. We carried out several classification experiments using the data sets described above. The first step of such an experiment is the generation of an all-against-all distance matrix, containing all scores of every pairwise comparison. Subsequently, either a *k*-leave-one-out or a 10-fold cross-validation is applied on the matrix.

In a *k*-leave-one-out cross-validation each line of the distance matrix is analyzed with respect to the *k* NN structures, and the query structure itself is omitted. The query is then assumed to be member of the same class to which the majority of the next NN belong. After evaluating all lines of the matrix, the total rate of correct classifications can be obtained by comparing the predicted classes of all query structures to the ones assigned by an independent method.

The 10-fold cross-validation is realized in a rather similar way. However, at the beginning of the process, the distance matrix is horizontally split into 10 equally large portions. Then, a leave-one-out cross-validation using only one NN ($k = 1$) is carried out for every single portion. Ten classification rates are obtained, which are in the following used to calculate the mean and standard deviation. This appears more reasonable than

dx.doi.org/10.1021/ci5005898 | J. Chem. Inf. Model. 2015, 55, 165−179

simply evaluating a single classification rate as it is returned in a k-leave-one-out cross-validation.

Inhibitors for the NMDA receptor. For the treatment of various neurological disorders, such as depression, Alzheimer's disease, or Parkinson's disease, new potent NMDA receptor inhibitors are searched.[69,70] Present investigations are focused on the detection of negative allosteric modulators (NAMs) rather than competitive ligands. NAMs may allow some kind of fine-tuning, whereas competitive ligands at the glutamate binding site can lead to complete channel blocking causing severe side effects. Ifenprodil is the prototype of modulatory compounds that bind at the NR1/NR2B interface of the amino terminal domain (ATD) of NMDA receptors.[70,71] Ifenprodil and similar derivatives are potent noncompetitive NMDA antagonists but also display a number of off-target activities, for example, interaction with adrenergic and sigma receptors and hERG blockade. Thus, there are many efforts to discover new compound classes with improved selectivity and similar or enhanced NMDA potency.[70]

In this context, we have applied our approach RAPMAD. Recently, crystal structures of GluN1/GluN2B ATDs in complex with ifenprodil and Ro 25-6981, another potent and selective NMDA blocker, have been deposited in the PDB[69,72,73] (PDB codes: 3QEM, 3QEL, 4PE5, 4TLL, 4TLM). Their pockets, automatically extracted by the LIGSITE algorithm, were compared against the entire Cavbase database and thus were used as the starting point of a virtual screening campaign in order to identify other potent inhibitors of the NMDA receptor. Ligands of the most similar pockets were used as a first idea for possible new inhibitors. With several filtering steps, such as docking and visual inspection, better insights were gained into the binding modes of the retrieved candidates, and the set could be further reduced to the most reasonable structures. Finally, the top-ranked and commercially available ligands were experimentally tested by determining IC_{50} values using an NMDA-binding assay developed for the binding of ifenprodil. A detailed protocol of the experimental procedure can be found in the Supporting Information.

RESULTS AND DISCUSSIONS

Classification Tests on Two-Class Data Sets. To evaluate our newly developed comparison method, we start with a classification experiment using the well-established ATP/NAD_{small} data set. This data set has already been the subject of many previous comparative studies to evaluate new cavity matching approaches with respect to the original Cavbase implementation. We therefore regard this data set as a reference standard benchmark for the evaluation of binding site comparison methods. Relying on the work of Fober et al.[36] and Mernberger et al.,[38] we can compare RAPMAD to a variety of other methods. In particular, we compare our method with the results obtained by the clique comparison algorithm originally implemented into Cavbase. Therefore, we first calculate the $n \times n$ distance matrix and apply a k-leave-one-out cross-validation subsequently. According to the results reported in the above-mentioned studies, we used different values for k ranging from 1 to 9. Remarkably, many algorithms perform better than the original Cavbase implementation, although they do not take any information about the local shape of the cavity surface into account. Table 2 lists the results of various comparison methods. Presented values indicate the percentage of correct classifications, when a varying number of nearest neighbors was used as classifier. The graph alignment approach (GA) was proposed by Weskamp et al.[74] GAVEO, SEGA, and the Bron-Kerbosch algorithm (BK) are also graph-based methods. LPCS is a geometrical approach in order to find optimal superpositions of labeled point clouds in 3D space. RAPMAD achieves higher classification rates than the original Cavbase implementation and many of the other methods presented previously. Only LPCS and SEGA achieve better results but require significantly longer runtime. The experiment is carried out by RAPMAD in around 5.6 s on a single core. On the contrary, SEGA and LPCS need 2.7 and 3.3 days instead, and Cavbase takes even more than 1 week to finalize. This comparison indicates already that the novel RAPMAD approach makes binding site comparisons feasible interactively on the fly. A more detailed analysis regarding the runtime will be demonstrated later in our study.

Table 2. Comparison of Classification Results of RAPMAD with Respect to the ATP/NAD_{small} Data Set Evaluated by Several Other Binding Site Comparison Methods, Including the Original Cavbase Implementation

k	Cavbase	GA	GAVEO	LPCS	SEGA	BK	RAPMAD
1	81.7	76.6	78.9	93.5	91.6	83.4	88.7
3	83.1	71.8	76.6	91.6	92.4	82.8	87.3
5	83.1	72.4	78.0	89.0	91.3	81.4	85.4
7	81.1	71.8	78.6	88.5	91.6	80.6	84.8
9	79.4	71.3	76.6	86.2	-	81.4	83.9

Because the coenzymes ATP and NAD vary clearly in size, it could well be that already a simple size criterion assigned to the pockets differs sufficiently and can be adequately captured by the histogram approach. To rule out the suspicion that convincing results of RAPMAD are solely based on pocket size differences with respect to their maximal diameter, we performed another classification experiment based on the same data set. This time the distance score has been determined solely based on the maximal diameter of a pocket as discrimination criterion. The maximal diameter can be easily obtained by calculating all pairwise distances of the pseudocenters in a pocket. The new distance score S_d was then defined as the difference of two maximal diameters D_{max}, which is calculated as

$$D_{max} = \max_{i=1\ldots|P|, j=1\ldots|P|} d(p_i, p_j)$$

where P represents the set of pseudocenters of the pocket, and $d(p_i,p_j)$ returns the Euclidean distance between the pseudocenters p_i and p_j. S_d is finally calculated as

$$S_d = |Da_{max} - Db_{max}|$$

where Da_{max} and Db_{max} represent the maximal distances in pockets A and B, respectively. Using this criterion, very poor results are obtained as shown in Table 3. Also, the average number of pseudocenters that comprises a binding pocket is unlikely a sufficiently reliable discriminating feature across the data set, as it amounts to 184.89 ± 168.94 for ATP and 144.73 ± 93.55 for NAD. Thus, the two distributions do not differ significantly. Even considering the number of type-specific pseudocenters is also not sufficient to discriminate the pockets successfully, as the correct classifications decrease to an average rate of 71.2% in this case. If neither pocket size nor the number of pseudocenters is the important information that is captured by the histograms, it must be coded in the spatial distribution of the pseudocenters.

dx.doi.org/10.1021/ci5005898 | J. Chem. Inf. Model. 2015, 55, 165−179

Table 3. Classification Results of RAPMAD on the ATP/NAD_{small} Data Set When Using the Difference between the Maximal Diameters of Two Pockets as a Measure of Dissimilarity

k	classification rate (%)
1	51.8
3	57.2
5	56.3
7	55.2
9	56.3

To evaluate the robustness of our method with respect to pocket flexibility or crystallographic errors or uncertainties (e.g., assignment of donor/acceptor properties in Asn, Gln, or His), we assessed the sensitivity of our comparison procedure by applying a random perturbation of the assigned pseudocenter positions. In detail, the x, y, and z coordinates of all pseudocenters were randomly modulated within local boundaries of −2 to +2 Å before calculating the histograms, which led to an average RMSD between the original and the perturbed pseudocenter sets of 1.99 Å. In fact, the classification rates were only deteriorated on average by 4.7% for k = 1...9, indicating significant robustness with respect to these perturbations.

To further evaluate the discriminative power of RAPMAD compared to the original Cavbase implementation, we focused next on the larger and more complex data set ATP/NAD_{large}. Moreover, we switched to the 10-fold cross-validation as the evaluation technique because it also provides the standard deviation of the classification rates as additional information.

As already mentioned in an earlier study, the Cavbase implementation achieves 79.2 ± 3.1% using this data set.[59] Using the BK algorithm to determine the maximum common subgraph of two graphs representing the binding sites to be compared achieves 84.6 ± 2.4%. Also, here LPCS and SEGA, which also provided very good classification results for ATP/NAD_{small}, perform best with 91.6 ± 2.2% and 94.3 ± 3.5%. Nevertheless, RAPMAD still excels BK and Cavbase, attaining 86.4 ± 3.2% (Figure 6a). It is important to note that the results obtained by LPCS and SEGA are based on individually optimized parameters for the given data sets.

We complete our classification tests considering the cofactor data sets that incorporated FAD binding sites (Figure 6b). In all cases, our novel implementation achieves considerably good results. The comparison between pockets of the data sets FAD and NAD_{large} was expected to be rather challenging due to the similarity in molecular structure of the two cofactors. Nevertheless, RAPMAD reaches an 85.1 ± 5.0% correct classification rate in this case. Pockets hosting FAD or ATP can be even differentiated more easily. Here, results of 88.7 ± 3.6% are attained. Before running the experiment that concerns the FAD and FAD_{cov} data sets, we randomly reduced the number of pockets in the FAD set from 429 to 114 in order to achieve two equally populated sets of pockets in both cases to avoid classification of a highly biased data set. Remarkably, the best results are obtained on this data sets, where the mean rate of correct classifications exceeds 90% (90.5 ± 5.9%). Also, in this case, the discriminating factor reflected by the shapes of the histograms must be coded in the spatial distribution of the pseudocenters because the pockets in the two data sets FAD and FAD_{cov} consist of an almost identical number of pseudocenters (170 ± 55 and 177 ± 45). In this case, we attribute the good results to the different overall shape of the binding sites in FAD_{cov}, as these structures tend to host the cofactor in a much more buried fashion.

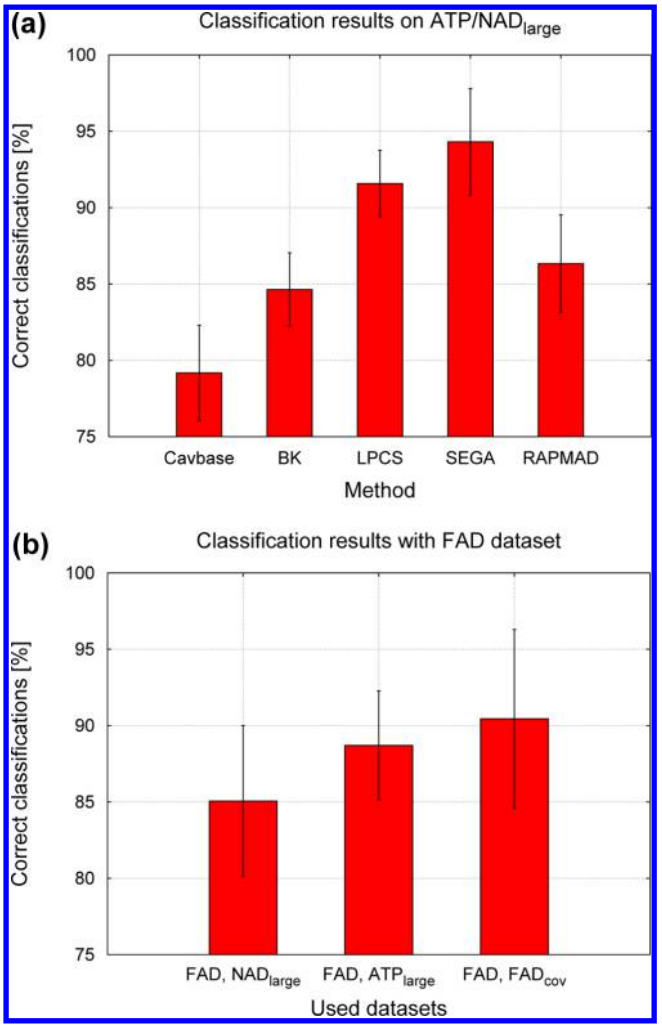


Figure 6. Classification results regarding the data sets ATP/NAD_{large}, FAD, and FAD_{cov}. (a) Classification results of different approaches for the ATP/NAD_{large} data set. Cavbase performs worst, leading to only 79.18 ± 3.12%. LPCS and SEGA perform almost equally well, attaining 91.58 ± 2.17% and 94.31 ± 3.48%, respectively. The BK algorithm returns considerable rates of 84.64 ± 2.40% as well as RAPMAD, which still achieves rather impressive rates of 86.34 ± 3.19%. (b) Discriminative power of RAPMAD by performing classification tests on the data sets FAD vs NAD_{large}, FAD vs ATP_{large}, and FAD vs FAD_{cov} (FAD includes only noncovalent pockets). In all experiments shown, our method leads to convincing results, especially when comparing the pockets binding FAD noncovalently to those which bind it covalently (90.45 ± 5.85%, right).

Classification Tests on a Multiple-Class Protease Data Set. The small protease data set, which is based on the Merops classification, was utilized for the next evaluation. It includes 84 pockets split into three clans, namely, serine, cysteine, and metalloproteases. For assessment purposes, we also calculated the scoring matrix by using the Cavbase implementation. As the portions would be way too small to apply a 10-fold cross-validation on this matrix, we used the k-leave-one-out cross-validation for evaluation. Table 4 shows the classification rates of this experiment when either RAPMAD or the comparison approach originally implemented in Cavbase is used. In the Table, the column RAPMAD* indicates the results after removing artificially added metal ions from the structures 1Q3X

dx.doi.org/10.1021/ci5005898 | J. Chem. Inf. Model. 2015, 55, 165−179

Table 4. Classification Results of Cavbase and RAPMAD on the Protease Dataset

k	Cavbase	RAPMAD	RAPMAD*
1	86.9	79.8	84.5
3	83.3	82.1	84.5
5	79.8	85.7	88.1
7	77.4	84.5	86.9
9	73.8	83.3	85.7

and 2QXJ. As classifier, the k-NN method was used with k = 1...9. Cavbase clearly deteriorates with increasing k, while the rates of RAPMAD improve. In general, however, the results of RAPMAD are more stable and do not vary as much as the Cavbase rates depending on the used classifier (standard deviation 2.3 vs 5.1). Moreover, Cavbase omits about 17% of the comparisons due to an overflow of its internal memory limitations. In RAPMAD, such problems are very unlikely and have never been observed in any of the trials executed in this study. As mentioned earlier, the most challenging classification when using the k-NN classifier is given when all the used classes comprise equal size. By randomly discarding seven entries from the metallo protease set and 20 examples from the serine proteases, we succeeded in obtaining three equal-sized data sets (19 pockets in each subset). For this experiment, the results do not change remarkably from the previous ones considering the larger and less balanced data set. They now range between 80.7% and 84.2% for k = 1...9, which indicates no substantial bias regarding the size differences of the used protease classes. Also, recalculating the classification rates only on the common subset (as Cavbase has omitted 17% of the comparisons) does not provoke remarkable changes of the results.

In addition, RAPMAD was able to detect two clear outliers in the serine protease set (see red arrows in Figure 7), as they exhibited metal ions in their catalytic sites. In one structure (PDB code 1Q3X), a sodium ion was found, which was picked up from the buffer medium during the crystallization. In the second structure (PDB code 2QXJ), two copper ions are present that have been experimentally added to investigate their enhancing or inhibiting influence on enzyme activity. After removal of the metal pseudocenters from the cavity descriptions of these two binding pockets, the accuracy of our method could be even further improved, which is shown in the column denoted by RAPMAD* (Table 4). On the contrary, the results obtained by the Cavbase implementation did not change remarkably. In the corresponding heat map of RAPMAD's scoring matrix (Figure 7), the regions of high similarity across the serine and cysteine proteases can be easily detected (upper right and lower left corner). Also, for the metalloproteases, a darker region can be identified, albeit less discriminating. The heat map suggests furthermore a close structural relationship between the active sites of cysteine and serine proteases, which is demonstrated off diagonal by the redish areas in the upper left and lower right corner. For both enzyme species, the mechanism follows very similar principles. It is known that serine proteases can be successfully morphed into cysteine proteases and vice versa by exchanging the catalytic cysteine and serine residues,[75] and there is also a Merops clan PA that contains serine proteases as well as cysteine proteases. Nevertheless, we observed that RAPMAD is able to obtain correct retrieval rates in more than 80% if only cysteine and serine proteases are subjected to a classification experiment.

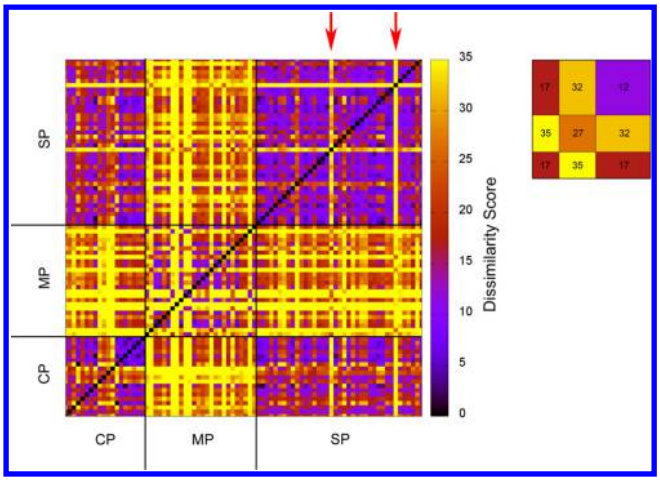


Figure 7. Heat map of the $n \times n$ dissimilarity matrix for the protease data set. Next to the detailed heat map a smaller version is pictured that shows the average distances scores of all important areas. Areas of high similarities (violet) are obtained for the comparisons of pockets assigned to the same class. In the upper right corner, the matrix serine proteases (SP) are compared with each other with a mean distance score of 12, and in the lower left corner are cysteine proteases (CP) with a mean similarity score of 17. The region where metalloproteases (MP) are compared with each other remains more fragmented (mean score of 27), but nevertheless it discriminates well from the other protease pockets. Off from the main diagonal, the detailed heat map also indicates structural similarity between SP and CP, which is highlighted by the darker regions in the lower right and upper left corner. Red arrows identify SP pockets from 1Q3X and 2QXJ that are obviously rather dissimilar to all the remaining SP structures.

Classification Tests on EC Data Sets. As the last classification experiment, we considered the EC and serine protease data sets. With respect to the EC data set, we succeeded in 94.2 ± 2.4% correct classifications. Thus, RAPMAD is apparently very successful in the differentiation of protein binding sites with respect to independently annotated EC numbers.

Nonetheless the EC data set is too small to allow for general conclusions. We thus decided to consider a larger serine protease data set, which consists of twice as many binding sites (in total 1028 entries). In the following experiment, we removed the pockets that have not been annotated to any specific EC number yet, namely, those artificially assigned to subclass "0" (65 entries) and, moreover, discarded all subclasses comprising only one pocket entry, as classification with the k-NN classifier requires at least two pockets. Subsequently, we created three subsets from the remaining subclasses for classification tests. In a first step, all 33 remaining subclasses were used, and RAPMAD achieves an assignment accuracy of 86.63 ± 2.94%. To avoid the above-mentioned problem of trivial majority voting, which results from classifying attempts based on imbalanced population, we decided to use in the next step only the classes comprising between 15 and 75 pockets. Ten classes remained, and the accuracy of RAPMAD improved to 92.37 ± 5.61%. Finally, only the two largest subclasses 4 and 5 (trypsin and thrombin), which include 275 and 225 pockets, respectively, were used to perform a two-class assignment experiment. In this case, our algorithm performs best, succeeding in 96.8 ± 1.69% correct classifications. These experiments demonstrate the discriminative power of our method to correctly assign even closely related proteins, such as

dx.doi.org/10.1021/ci5005898 | J. Chem. Inf. Model. 2015, 55, 165−179

binding sites of serine proteases of the trypsin family that only differ in the last digit of the EC number.

We were also interested in which EC classes RAPMAD assigns the above-mentioned entries of subclass "0". A listing of all 65 pockets together with the most similar structures of the remaining data set shows a remarkably high success rate (Supporting Information). In many cases, pockets of either the same proteins or proteins with a very high sequence similarity were matched (nine and 11 cases, respectively). In 10 cases, our method furthermore discovered the true EC class of elements that were so far assigned to subclass 0. Here, the most similar pocket affiliated to a different subclass than 0, and a subsequent consulting of the PDB confirmed that also the query pocket actually belongs to the same subclass. These findings are even more considerable as all the relationships were detected by solely using putative binding sites, which are usually quite small portions of the entire protein surfaces.

Retrieval of Enzymes Catalyzing the Same Reactions. To assess the retrieval power of RAPMAD, we tried to discover cavities of enzymes that catalyze the same chemical reactions. This search was based on the assumption that enzymes catalyzing the same reactions will share 3D similarity of their catalytic centers. We have chosen a binding pocket of a carbonic anhydrase II (EC number 4.2.1.1) as a query structure (PDB code 1OQ5) and matched this entry against all cavities with EC annotation in the cavity database. Thereby, a retrospective functional annotation study was conducted. This resulted in a total number of 161 822 comparisons, including 801 cavities assigned to the same EC annotation as the query. These structures were defined as correct hits. To face the RAPMAD comparison with the former method, we also evaluated this query with the original Cavbase comparison procedure. The results of both retrievals are illustrated by the ROC curves in Figure 8. At the very beginning, Cavbase achieves a slightly better enrichment than RAPMAD, approaching an enrichment factor at 1% (EF 1%) of 0.46, where RAPMAD reaches 0.41. But, after a comparison of about 7% of the database, the results of RAPMAD outperform Cavbase and lead to a higher AUC of 0.75, whereas Cavbase achieves 0.69. This difference has also to be assessed with respect to the runtimes of both approaches. When RAPMAD has scanned the entire database, the Cavbase implementation has virtually processed less than 0.01% of all entries.

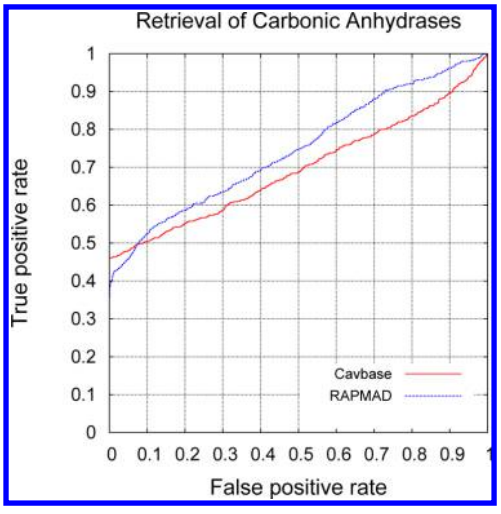


Figure 8. ROC curves illustrating the retrieval results of carbonic anhydrase pockets unsing RAPMAD and Cavbase. A binding site of a carbonic anhydrase (PDB code 1OQ5, EC number 4.2.1.1) was used as query and compared against all EC-annotated pockets in Cavbase. The plots illustrate the detection of cavities with the same EC number as the query structure.

Similar experiments were conducted using a trypsin and a subtilisin pocket as reference for a protease query. As pointed out in ref 25, trypsin and subtilisin exhibit the same biochemical function by executing the same enzyme mechanism, but they share no sequence or fold homology. We used as query the pocket of 1TPO for trypsin and 2PRK for subtilisin. In a first figure-of-merit, the performance of retrieving pockets that belong to the same family as the query pocket was evaluated. The retrieval based on trypsin as reference pocket achieves both for Cavbase and RAPMAD the same AUC of 0.68. In the case of subtilisin as the reference, Cavbase performs slightly better, reaching an AUC of 0.75, whereas RAPMAD succeeds in 0.71. Subsequently, we removed all pockets of the same family apart from the query from the list of results and analyzed the performance of the pocket used for retrieval with respect to the other class. While using the trypsin pocket as the query to retrieve the subtilisin example, Cavbase and RAPMAD attained comparable AUCs of 0.69 and 0.68, respectively. Using the subtilisin pocket to extract trypsin binding sites, the retrieval even resulted in identical AUCs of 0.65 for Cavbase and RAPMAD. Taking all five experiments together, RAPMAD achieves on average an AUC of 0.694, while Cavbase achieves 0.692, which demonstrates once more equal performance of RAPMAD compared to the currently implemented approach in Cavbase. The results are summarized in Table 5.

Retrieval of Proteins Binding a Particular Ligand. In the next step, we will focus on binding sites present in different proteins that bind the same ligand. This time we will focus on pharmaceutically relevant drugs that are present in a fairly large set of PDB entries. As a first example, we performed a retrieval experiment on pockets that bind methotrexate, a chemotherapeutic drug, which is also used for the treatment of autoimmune diseases. By searching the PDB, we found 30 structures that bind methotrexate and are also stored in our current version of Cavbase. We then used the protein culling server PISCES[76] to reduce the putatively given redundancy in sequence space of this data set. The sequence identity threshold was set to 20% in order to consider only proteins beyond the so-called twilight zone threshold with respect to sequence.[77] The minimal resolution was asked to be below 2.5 Å, and the highest acceptable R-factor was 30%, which resulted in four remaining structures: a pteridine reductase (1E7W), a thymidylate synthase (1AXW), and two dihydrofolate reductases (3DFR, 3DAU). The flexible structure comparison method FATCAT[78] subsequently returned that all selected structures do not exhibit significant similarity except the dihydrofolate reductases 3DFR and 3DAU, which motivated us to discard one of these two structures. We decided to remove the entry with lower resolution. Thus, we finally considered the three structures 1E7W, 1AXW, and 3DAU. Next, we consulted Cavbase to retrieve all pockets accommodating methotrexate in these proteins, which suggested five pockets overall because we found two independent binding pockets for 1E7W and 1AXW. These binding pockets also accommodate in addition the cofactor NADPH dihydronicotinamide-adenine-dinucleotide phosphate or 2′-deoxyuridine 5′-monophosphate, respectively. Nonetheless, all comparisons were performed based on the entire binding

dx.doi.org/10.1021/ci5005898 | J. Chem. Inf. Model. 2015, 55, 165−179

Table 5. Overview of Retrieval Results To Detect Similar Binding Sites Validated in Accordance to the Annotated EC Number

task	AUC Cavbase	AUC RAPMAD
retrieving carbonic anhydrases using 1OQ5.1	0.69	0.75
retrieving trypsins using 1TPO.1	0.68	0.68
retrieving subtilisins using 2PRK.1	0.75	0.71
retrieving subtilisins using 1TPO.1	0.69	0.68
retrieving trypsins using 2PRK.1	0.65	0.65
average AUC	0.69	0.69

pocket instead of limiting solely to the regions hosting methotrexate. Thereby, we can assess the capability of RAPMAD to deal with cavities that are significantly larger than the actual ligand-binding region of interest. For any of the pockets, we performed mutual comparisons against the entire cavity database and plotted ROC curves showing at which stage the remaining methotrexate pockets were retrieved (Figure 9).

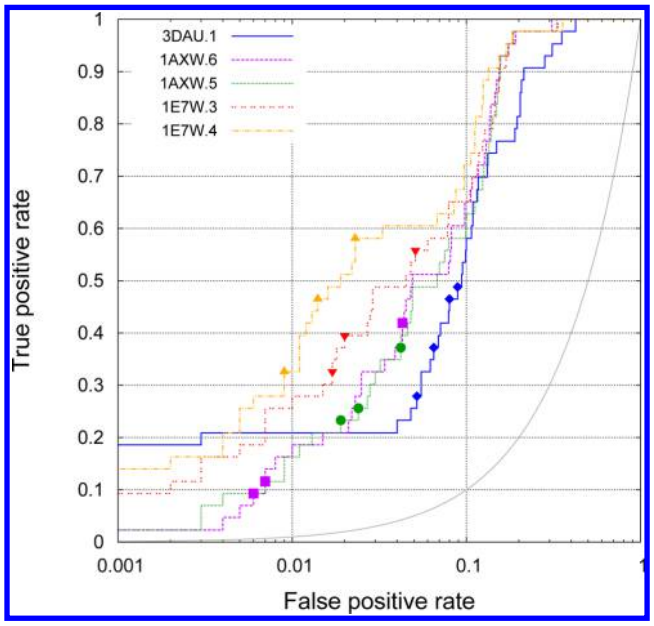


Figure 9. Logarithmically scaled ROC curves for the retrieval of protein binding sites that host the drug methotrexate. Five pockets of the sequentially and structurally unrelated proteins 3DAU, 1AXW, and 1E7W were used as queries and compared against the entire database. To plot the ROC curves, a set of 43 additional methotrexate-binding pockets have been defined as a hit. In addition, the detection levels of the pockets corresponding to the other two unrelated proteins are highlighted as triangles, circles, and rectangles. A random performance is indicated by the gray line. All ROC curves exhibit an average AUC of 0.92.

To define a hit set, we retrieved all cavities from Cavbase that actually accommodate methotrexate (50 in total). By visual inspection, six structures were subsequently omitted because LIGSITE was obviously not able to detect the methotrexate binding site properly (e.g., volume of more than 4000 Å^3 was assigned or two independent binding sites were joined), which finally resulted in 44 pockets.

In the following, we paid special attention to the pockets of the sequentially and structurally unrelated proteins, which are marked as triangles, circles, and rectangles in Figure 9. As shown in the ROC plots, RAPMAD has no difficulties to detect pockets originating from the unrelated proteins because they are all discovered among the first 10% of the retrieval list (marked by 0.1 on the *x*-axis). This result appears very satisfactory, as it underlines that pocket similarity is only defined by the position of the pseudocenters and does not result from inherent sequence or structural folding similarity. Moreover, the ROC curves show that independent of the actually selected query pocket at least half of the stored methotrexate-binding pockets are retrieved among the first 10%, which indicates a rather promising early enrichment. The curves also exhibit a very convincing average AUC of 0.92.

As a second example, we used the binding pockets that host pemetrexed, a chemotherapy drug for the treatment of pleural mesothelioma and cell lung cancer. Again, we searched the PDB for proteins that bind pemetrexed and found six structures: four thymidylate synthases (1JUJ, 1JU6, 3K2H, 4FQS), a folate receptor (4KN2), and a pteridine reductase (2X9G). By culling this set of proteins using PISCES in order to reduce sequential redundancy, it was limited to three entries (3K2H, 4FQS, 2X9G), and by applying FATCAT, we ended up with the two structures 2X9G and 4FQS. Next, the cavity database was exploited to identify the binding sites of these proteins that actually accommodate the desired ligand. In doing so, four binding sites for 2X9G and two for 4FQS could be retrieved from the database. The hit set of pemetrexed-binding sites that could be found in Cavbase comprised 19 pockets. The resulting ROC curves are shown in Figure 10, where detections of pockets that correspond to sequentially and structurally unrelated proteins are particularly highlighted. Also, for this example, the ROC curves indicate that pockets of unrelated proteins are retrieved on very early ranks. For example, the first pteridine reductase example is on average detected after 1.2% of the data set and the first thymidylate synthase even after 0.17%! The curves in this experiment achieve an even higher average AUC, which is 0.96 and clearly demonstrates that RAPMAD can be used as a tool to seek for putative cross-reactivity.

Identification of New Ligands for the NMDA Receptor. Stimulated by the last example, we wanted to apply RAPMAD to a predictive case study. Our goal was the identification of potential NAMs for the NMDA receptor. The search was initiated with a retrieval for the most similar pockets to the ifenprodil-binding sites of 3QEL and 3QEM. As these are homodimers, LIGSITE detected two pockets for ifenprodil (Figure 11a) in each of the structures. In order to limit background noise and to focus on the essential information in the search process, the four pockets were reduced to the pseudocenters in an area of 8 Å around the ligand prior to the subjection to comparative searches. Next, we used the methods RAPMAD and Local Cliques (LC), a graph-based approach, which we presented in a previous study,[59] to compare the four query pockets against all entries in the cavity database. These individual screening runs revealed eight ranking lists. In each case, the top 250 entries were extracted and merged to one single hit list, and subsequently, any duplicate pocket entries were removed. For the remaining 1562 pockets, we extracted the accommodated ligands and reduced this collection further

dx.doi.org/10.1021/ci5005898 | J. Chem. Inf. Model. 2015, 55, 165−179

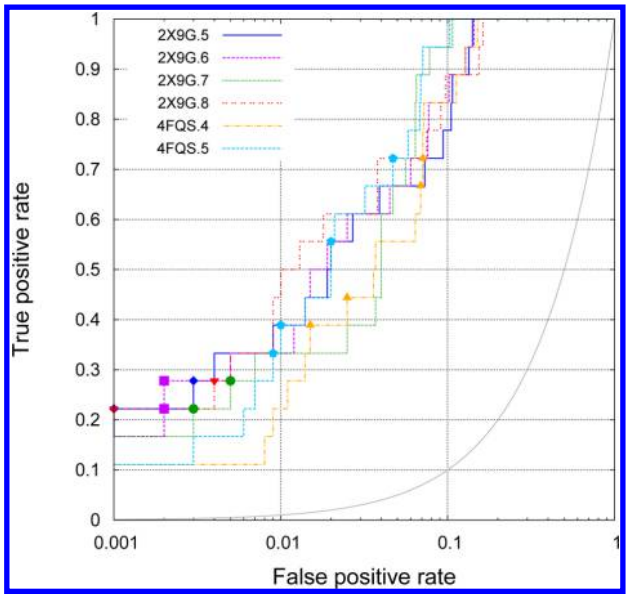


Figure 10. Logarithmically scaled ROC curves for the retrieval of protein binding sites that host the drug pemetrexed. Six pockets of the sequentially and structurally unrelated proteins 2X9G and 4FQS were used as queries and compared against the entire database. To plot the ROC curves, a set of 18 other pemetrexed-binding pockets have been defined as a hit. In addition, the detection levels of the pockets corresponding to the other unrelated protein are highlighted as triangles, circles, and rectangles. A random performance is indicated by the gray line. The ROC curves exhibit an average AUC of 0.96.

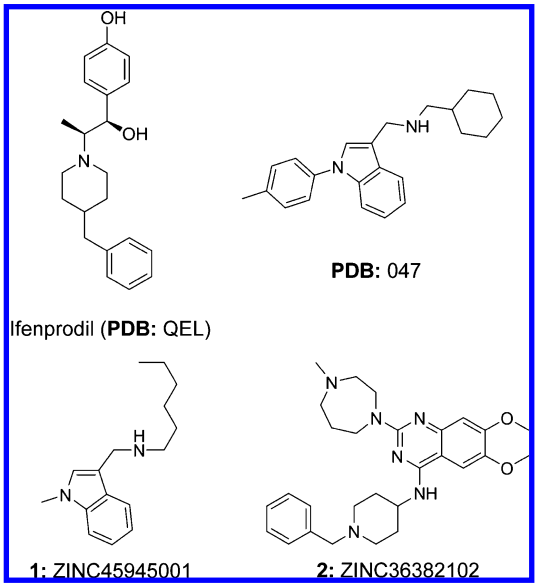


Figure 11. Results of the virtual screening for novel NMDA receptor inhibitors. (a) The negative allosteric modulator ifenprodil. (b) Ligand structure with PDB identifier 047. This molecule could not be ordered from ZINC. (c) Hits of the binding assay that exhibit substantial affinity to the NMDA receptor. Compound 1 shows an IC_{50} of 1051 nM and compound 2 an IC_{50} of 1560 nM.

by visual inspection. Molecules that did not match with our predefined criteria regarding size and druglikeness were discarded, resulting in a set of 97 remaining ligands. Subsequently, the ligands were docked into the target pocket of 3QEL using the program GOLD, which yielded 16 promising hits. Of these, 15 structures—or at least similar derivatives—were referenced in the ZINC database, and finally, we purchased five compounds for testing in a binding assay. The remaining hits were already present in our in-house database of tested NMDA binders; thus, no repetitive test was required. In case an exact match of the retrieved structure was not possible in the ZINC database, the most similar structure was ordered.

Two of the tested compounds showed significant binding affinity to the NMDA receptor (Figure 11c). Compound 1 exhibits an IC_{50} of 1051 nM. It is a similar compound to the ligand with PDB identifier 047 (Figure 11b) as the latter could not be ordered from ZINC. In the initial screening step (pocket comparisons), the binding site hosting this ligand was found by the method LC on rank 51, which is within the top 0.019% of the ranking list, when the pocket 3QEM.6 was used as the query. Compound 2 achieves an IC_{50} of 1560 nM and was found by RAPMAD on position 133 (top 0.048%) using the pocket 3QEL.8 as the query. In a follow-up step, the hit compounds could possibly be modified synthetically by decorating them with additional functional groups, or specific moieties could be substituted by bioisosteric portions to further increase their binding affinities. This follow-up study will afford additional resources in synthesis, but this would be by far out of the scope of this feasibility study. Nonetheless, this predictive example shows that RAPMAD can be successfully applied to retrieve new ligands via a comparative binding pocket approach.

Runtimes. Major superiority of RAPMAD compared to all previously developed methods is its extremely short runtime without sacrificing any accuracy. The explanation for this impressive speed-up is the sole evaluation of distance histograms that are based upon certain reference points and can both be constructed and compared with linear complexity $O(n)$. This is an utmost improvement because the fastest hitherto known methods work with at least cubic complexity $O(n^3)$. The runtime of RAPMAD will now be analyzed in detail.

Building Histograms. Let n be the number of pseudocenters in a protein binding site. At first, seven subsets are generated from the total set of pseudocenters, which requires n calculations. After that, the coordinates of the centroid and the centroid closest are calculated for each subset ($2n$ calculations in total). Finally, for each of the seven subsets, the distances from the centroid and the centroid closest to all pseudocenters are measured, which requires $7 \times 2 \times n$ calculations. In sum, $n + 2n + 7 \times 2 \times n = 17n$ calculation steps are performed, which is written as $O(n)$.

Comparison of Pockets. Let $m = \max(|H_A|, |H_B|)$ be the maximum number of bins of two histograms that are used in a pairwise comparison. For all the seven subsets of pseudocenters, the mean distribution M of the two input histograms has to be computed (m calculations). Subsequently, the Kullback–Leibler divergences KLD_A and KLD_B are determined in $2m$ steps. Thus, the comparison of two protein binding sites requires $7(m + 2m) = 21m$ calculation steps, which is again in linear complexity $O(m)$.

To demonstrate the actually achieved speed-up, we compared the runtime of RAPMAD to those of the previously developed programs. The runtimes of GA and BK were adopted from previous studies.[38,79] While determining the runtimes of the original Cavbase implementation, LPCS, SEGA, and RAPMAD, we used a state-of-the-art computer equipped with an Intel Core i7-3770 CPU (3.4 GHz) processor and 16

dx.doi.org/10.1021/ci5005898 | J. Chem. Inf. Model. 2015, 55, 165−179

Table 6. Mean Runtimes in Seconds and Standard Deviations of Several Binding Site Comparison Approaches and RAPMAD Using 1000 Randomly Chosen Pairs of Binding Sites for Retrieval

Cavbase	GA	GAVEO	LPCS	SEGA	BK	RAPMAD
12.76 ± 44.90	121.74 ± 418.02	584.51 ± 2199.02	2.24 ± 1.27	1.85 ± 2.27	6.91 ± 32.59	0.000045 ± 0.00012

GB memory. All times have been measured using only a single core of the device and a benchmark set of 1000 randomly chosen pairs of protein binding pockets were compared. Table 6 lists the mean runtimes and the standard deviations. Obviously, RAPMAD requires exceptionally short runtimes that are more than 40 000 times faster than those of LPCS or SEGA (all Java implementations), while sacrifying only a minor amount of accuracy with respect to the classification. Most important, the currently implemented binding site comparison approach in Cavbase (C++ implementation) is outperformed by more than 5 orders of magnitude. Figure 12 illustrates the

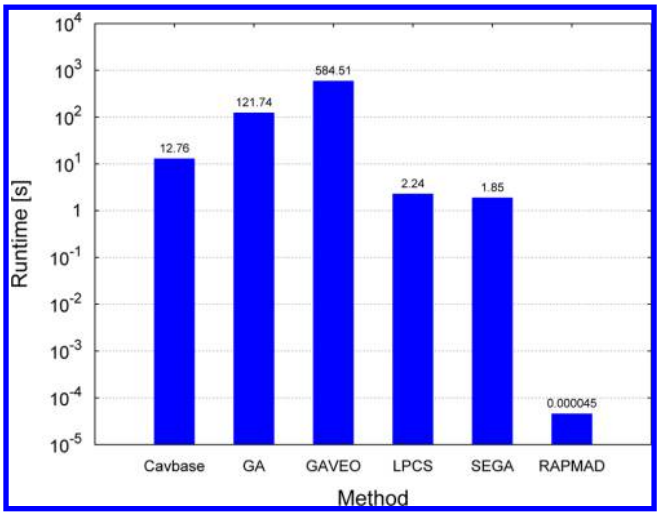


Figure 12. Average runtimes in seconds that are required for a single binding site comparison by all previously developed methods. The runtime is plotted on the *y*-axis in logarithmic scale. The ultrafast RAPMAD clearly sticks out because it is more than 4 orders of magnitude faster than the fastest hitherto developed algorithms. The algorithm presently implemented in Cavbase is outperformed by more than 5 orders of magnitude.

required runtimes on logarithmic(!) scale and stresses the magnitude difference. The diagram points out that a comparison of one pocket against the entire database (about 275 000 pockets) requires less than 13 s by using RAPMAD. The next faster methods SEGA and LPCS yet require 5.9 and 7.1 days, respectively, to fulfill this task. This advantage clearly moves binding pocket comparisons with RAPMAD into the scope of interactive modeling.

CONCLUSIONS

The detection of similar binding sites along with their occupants can greatly assist the drug development process particularly by suggesting bioisosteric molecular portions or novel scaffolds for optimization and predicting putative sites of cross-reactivity. In the present study, we describe the development, evaluation, and application of the novel similarity retrieval engine RAPMAD to detect putative protein binding sites, which are available from the Cavbase database. Our approach represents binding sites in terms of a set of translationally and rotationally invariant histograms. Undoubtedly, on first sight, the histogram representation implies a loss of information about the cavity shape, as no surface points are considered. But as shown in previous studies, most of the more recently developed comparative matching algorithms of Cavbase entries also operate without regarding surface points.[36−38] They solely consider pseudocenters and achieve impressive retrieval rates. Our new method, RAPMAD, accomplishs results of similar quality in both classification and retrieval that are comparable to the achievements of other state-of-the-art methods, however, important enough, in a fraction of time. It returns a distance score that can be interpreted as a measure of dissimilarity subsequently used to rank the compared pockets with respect to the query. Although a pass/fail decision whether two pockets are similar or not may be desirable in some cases, a graded ranking can be beneficial for other applications. However, this feature is shared by RAPMAD with the most of the other binding site comparison methods, which all are based on implemented scoring functions.

In classification tests on data sets of pockets that bind a specific cofactor, our procedure achieves substantially better results than the originally implemented algorithm in Cavbase. We showed furthermore that different types of proteases can be successfully discriminated with respect to a totally independent Merops classification. A set of serine proteases can be correctly classified with respect to the fourth digit of the assigned Enzyme Commission number. Also, enrichment experiments confirm the accuracy by retrieving correctly the active sites of target proteins on first ranks. Most importantly, all comparisons could be performed with hardly any loss of accuracy but at higher speed of more than 5 orders of magnitude compared to the currently implemented algorithm in Cavbase. When comparing the runtime of RAPMAD to the information published by the authors of other recently developed fast comparison methods, the efficiency of our appraoch can be underlined by the following figures. In one second and using a single CPU core, RAPMAD is able to perform about 20 000 pocket comparisons and thus clearly outperforms methods such as BSSF, FuzCav, PocketMatch, and TrixP.

Large-scale studies such as screenings of the entire Cavbase database, which comprises almost half a million cavities in its latest version (August 2014), become feasible using RAPMAD within seconds, making the tool applicable as an interactive idea generator. This was demonstrated by 11 database screenings that were carried out for query pockets binding to either the cancer drug pemetrexed or methotrexate. The results of these retrievals are convincing and stress that our method displays a valuable and extremely fast alternative to current binding site comparison procedures.

In our study, we focused on comparison procedures for putative protein binding sites that use the internal Cavbase format and descriptors. We appreciate that other powerful descriptor schemes to detect and compare pockets are available (such as FuzCav, PocketMatch, or BSSF); however, a direct 1:1 comparison of RAPMAD with these tools will be difficult as they built on distinct descriptor sets and cavity extraction algorithms. Apart from the search performance in terms of

dx.doi.org/10.1021/ci5005898 | J. Chem. Inf. Model. 2015, 55, 165−179

speed and retrieval rate, the success and relevance of such methods highly relies on the strategy applied to define similarity. Very different concepts have been used in the literature but comparing them would address a completely different aspect, which is not intended by this study.

Combining the presented method with other approaches for the comparison of Cavbase entries could help to further improve the accuracy by implementing a kind of consensus scoring. There is a variety of ways to realize such a combination (rank-by-number, rank-by-rank, rank-by-vote[80]), and it has been shown that consensus scoring in general delivers more robust results than a single method.[81] A combination of two or more highly efficient methods applied subsequently after the histogram approach will still be a lot faster than a single scoring as used in the current Cavbase implementation. In our study, we implemented a consensus scoring comprising RAPMAD and the more elaborate graph-based LC method. By using this combination as a virtual screening and data mining tool, we were able to identify promising hits in a binding assay regarding the affinity to the NMDA receptor. Testing five compounds revealed two ligands that exhibit IC_{50} values of about 1000 and 1500 nM, which corresponds to a hit rate of 40%.

Although our method leads to very good results in general, limitations become apparent when two binding sites are matched only in a certain region. RAPMAD estimates global similarity between binding sites and runs currently into problems when pockets hosting the same ligand deviate rather strongly in size and shape. This became obvious in the screenings for methotrexate-hosting pockets, where pockets hosting methotrexate only are found less successfully than those accommodating methotrexate along with a cofactor such as NAD, NDP, or UMP in the dehydrogenases or synthases.

The screening for binding pockets along with the retrieval of bound occupants as an idea generator is a typical modeling approach in drug design where a sketchy defined search query has to be fitted and optimized according to the achieved results. If the database screen requires however days to be accomplished, any interactive optimization of the search query is impossible. With RAPMAD, this very typical and appropriate working strategy becomes feasible.

Finally, our method is not restricted to protein binding sites alone but can easily be exploited for the comparison of other labeled point sets in 3D space, for example, pharmacophoric points across ligands. With regard to its linear complexity and ultrafast runtime, it might also be applied easily to much larger point sets, which will be impossible for graph-based methods due to their exceptionally long runtimes and high memory requirements. Further examinations could focus on exploring the diverse areas of applications.

■ ASSOCIATED CONTENT

Supporting Information

PDB codes and Cavbase cavity identifiers of all the protein binding pockets used in the data sets, protocol of the binding assay concerning the NMDA receptor, and a list of query pockets out of EC class 3.4.21."0," along with the most similar matches of the remaining serine protease data set. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: Klebe@staff.uni-marburg.de.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors gratefully acknowledge Dimitri Grigorev (Univ. Marburg), who took great efforts in compiling the new NAD, ATP, and FAD datasets. Thanks also go to Dr. Serghei Glinca (Univ. Marburg), who kindly provided us with the EC and protease datasets, and to Thomas Rickmeyer (Univ. Marburg), who ran several database screenings using the Cavbase system. Dr. Marc Strickert (Univ. Marburg) delivered us the serine protease dataset, and Dr. Thomas Fober (Univ. Marburg) provided us with some newly calculated results and runtimes of LPCS and SEGA. We furthermore thank Simon Müller, Bastian Weißkopf, and Benjamin Wenzel (Univ. Marburg) for carrying out several experiments and exhaustively testing the program.

■ REFERENCES

(1) Sim, S. C.; Ingelman-Sundberg, M. *Trends Pharmacol. Sci.* **2011**, *32*, 72−81.
(2) Mestres, J.; Gregori-Puigjané, E.; Valverde, S.; Solé, R. V. *Mol. BioSyst.* **2009**, *5*, 1051−1057.
(3) Nisius, B.; Sha, F.; Gohlke, H. *J. Biotechnol.* **2012**, *159*, 123−134.
(4) Kufareva, I.; Ilatovskiy, A. V.; Abagyan, R. *Nucleic Acids Res.* **2012**, *40*, D535−D540.
(5) Pérot, S.; Sperandio, O.; Miteva, M. A.; Camproux, A. C.; Villoutreix, B. O. *Drug Discovery Today* **2010**, *15*, 656−667.
(6) Schmidtke, P.; Souaille, C.; Estienne, F.; Baurin, N.; Kroemer, R. T. *J. Chem. Inf. Model.* **2010**, *50*, 2191−2200.
(7) Ghersi, D.; Sanchez, R. *J. Struct. Funct. Genomics* **2011**, *12*, 109−117.
(8) Thornton, J. M. *Science* **2001**, *292*, 2095−2097.
(9) Holm, L.; Sander, C. *Nucleic Acids Res.* **1997**, *25*, 231−234.
(10) Guda, C.; Lu, S.; Scheeff, E.; Bourne, P.; Shindyalov, I. *Nucleic Acids Res.* **2004**, *32*, W100−W103.
(11) Teichert, F.; Bastolla, U.; Porto, M. *BMC Bioinf.* **2007**, *8*, 425−442.
(12) Redfern, O.; Harrison, A.; Dallman, T.; Pearl, F.; Orengo, C. *PLoS Comput. Biol.* **2007**, *3*, 2334−2347.
(13) Murzin, A.; Brenner, S.; Hubbard, T.; Chothia, C. *J. Mol. Biol.* **1995**, *247*, 536−540.
(14) Orengo, C.; Michie, A.; Jones, S.; Jones, D.; Swindells, M.; Thornton, J. *Structure* **1997**, *5*, 1093−1108.
(15) Holm, L.; Sander, C. *Nucleic Acids Res.* **1996**, *24*, 206−209.
(16) Stark, A.; Russell, R. *Nucleic Acids Res.* **2003**, *31*, 3341−3344.
(17) Barker, J.; Thornton, J. *Bioinformatics* **2003**, *19*, 1644−1649.
(18) Moll, M.; Kavraki, L. *Matching of Structural Motifs Using Hashing on Residue Labels and Geometric Filtering for Protein Function Prediction*, CSB'08: 7th Conference on Computational Systems Bioinformatics. Proceedings, Palo Alto, CA, 2008; pp 157−169.
(19) Zauhar, R. J.; Moyna, G.; Tian, L.; Li, Z.; Welsh, W. J. *J. Med. Chem.* **2003**, *46*, 5674−5690.
(20) Binkowski, T.; Joachimiak, A. *BMC Structural Biology* **2008**, *8*, 45.
(21) Binkowski, T.; Adamian, L.; Liang, J. *J. Mol. Biol.* **2003**, *332*, 505−526.
(22) Kinoshita, K.; Nakamura, H. *Protein Sci.* **2003**, *12*, 1589−1595.
(23) Shulman-Peleg, A.; Nussinov, R.; Wolfson, H. J. *J. Mol. Biol.* **2004**, *339*, 607−633.
(24) Schmitt, S.; Hendlich, M.; Klebe, G. *Angew. Chem., Int. Ed.* **2001**, *40*, 3141−3146.
(25) Schmitt, S.; Kuhn, D.; Klebe, G. *J. Mol. Biol.* **2002**, *323*, 387−406.
(26) Weskamp, N.; Hüllermeier, E.; Klebe, G. *Proteins: Struct., Funct., Bioinf.* **2009**, *76*, 317−330.
(27) Hendlich, M.; Bergner, A.; Günther, J.; Klebe, G. *J. Mol. Biol.* **2003**, *326*, 607−620.

(28) Günther, J.; Bergner, A.; Hendlich, M.; Klebe, G. *J. Mol. Biol.* **2003**, *326*, 621−636.
(29) Hendlich, M.; Rippmann, F.; Barnickel, G. *J. Mol. Graphics Modell.* **1997**, *15*, 359−363.
(30) Berman, H.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.; Weissig, H.; Shindyalov, I.; Bourne, P. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(31) PyMOL. Schrödinger, LLC. http://www.schrodinger.com/pymol/ (accessed December 2014).
(32) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *J. Mol. Biol.* **1990**, *215*, 403−10.
(33) Chalk, A. J.; Worth, C. L.; Overington, J. P.; Chan, A. W. *J. Med. Chem.* **2004**, *47*, 3807−16.
(34) Rosen, M.; Lin, S. L.; Wolfson, H.; Nussinov, R. *Protein Eng.* **1998**, *11*, 263−77.
(35) Bron, C.; Kerbosch, J. *Commun. ACM* **1973**, *16*, 575−577.
(36) Fober, T.; Mernberger, M.; Klebe, G.; Hüllermeier, E. *Bioinformatics* **2009**, *25*, 2110−2117.
(37) Fober, T.; Glinca, S.; Klebe, G.; Hüllermeier, E. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2011**, *8*, 1653−1666.
(38) Mernberger, M.; Klebe, G.; Hüllermeier, E. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2011**, PrePrints.
(39) von Behren, M. M.; Volkamer, A.; Henzler, A. M.; Schomburg, K. T.; Urbaczek, S.; Rarey, M. *J. Chem. Inf. Model.* **2013**, *53*, 411−422.
(40) Volkamer, A.; Griewel, A.; Grombacher, T.; Rarey, M. *J. Chem. Inf. Model.* **2010**, *50*, 2041−2052.
(41) Yeturu, K.; Chandra, N. *BMC Bioinf.* **2008**, *9*, 543.
(42) Xiong, B.; Wu, J.; Burk, D. L.; Xue, M.; Jiang, H.; Shen, J. *BMC Bioinf.* **2010**, *11*, 47.
(43) Weill, N.; Rognan, D. *J. Chem. Inf. Model.* **2010**, *50*, 123−135.
(44) Meslamani, J.; Rognan, D.; Kellenberger, E. *Bioinformatics* **2011**, *27*, 1324−1326.
(45) Kahraman, A.; Morris, R.; Laskowski, R.; Thornton, J. *J. Mol. Biol.* **2007**, *368*, 283−301.
(46) Ballester, P. J.; Richards, W. G. *J. Comput. Chem.* **2007**, *28*, 1711−1723.
(47) Osada, R.; Funkhouser, T.; Chazelle, B.; Dobkin, D. *ACM Trans. Graphics* **2002**, *21*, 807−832.
(48) Schneider, G.; Baringhaus, K. H. *Molecular Design*; Wiley-VCH: Weinheim, Germany, 2008.
(49) Amorim, R. C.; Mirkin, B. *Pattern Recognit.* **2012**, *45*, 1061−1075.
(50) Kolmogorov, A. *Giornale dell'Istituto Italiano degli Attuari* **1933**, *4*, 83−91.
(51) Smirnov, N. *Ann. Math. Stat.* **1948**, *19*, 279−281.
(52) Jaccard, P. *Bull. Soc. Vaudoise Sci. Nat.* **1901**, *37*, 547−579.
(53) Hellinger, E. *J. Reine Angew. Math.* **1909**, *136*, 210−271.
(54) Lance, G. N.; Williams, W. T. *Aust. Comput. J.* **1967**, *1*, 15−20.
(55) Dagan, I.; Lee, L.; Pereira, F. *Similarity-Based Methods for Word Sense Disambiguation*, Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics, Stroudsburg, PA, U.S.A., 1997; pp 56−63.
(56) Kullback, S.; Leibler, R. A. *Ann. Math. Stat.* **1951**, *22*, 79−86.
(57) Neudert, G.; Klebe, G. *J. Chem. Inf. Model.* **2011**, *51*, 2731−2745.
(58) Gohlke, H.; Hendlich, M.; Klebe, G. *J. Mol. Biol.* **2000**, *2*, 337−356.
(59) Krotzky, T.; Fober, T.; Hüllermeier, E.; Klebe, G. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2014**, *11*, 878−890.
(60) Siggelkow, S.; Burkhardt, H. *Improvement of Histogram-Based Image Retrieval and Classification*, Proceedings of the 16th International Conference on Pattern Recognition (ICPR'02), Volume 3, Washington, DC, U.S.A., 2002; p 30367.
(61) Stegemann, B.; Klebe, G. *Proteins* **2011**, *80*, 626−648.
(62) Dobson, P. D.; Doig, A. J. *J. Mol. Biol.* **2005**, *345*, 187−199.
(63) Dym, O.; Eisenberg, D. *Protein Sci.* **2001**, *10*, 1712−1728.
(64) Glinca, S.; Klebe, G. *J. Chem. Inf. Model.* **2013**, *53*, 2082−2092.
(65) Schomburg, K. T.; Rarey, M. *J. Chem. Inf. Model.* **2014**, *54*, 2261−2274.
(66) Rawlings, N. D.; Barrett, A. J.; Bateman, A. *Nucleic Acids Res.* **2012**, *40*, D343−D350.
(67) Hanley, J. A.; McNeil, B. J. *Radiology* **1982**, *143*, 29−36.
(68) Bradley, A. P. *Pattern Recognit.* **1997**, *30*, 1145−1159.
(69) Karakas, E.; Simorowski, N.; Furukawa, H. *Nature* **2011**, *475*, 249−253.
(70) Mony, L.; Kew, J. N.; Gunthorpe, M. J.; Paoletti, P. *Br. J. Pharmacol.* **2009**, *157*, 1301−1317.
(71) Carter, C.; Benavides, J.; Legendre, P.; Vincent, J. D.; Noel, F.; Thuret, F.; Lloyd, K. G.; Arbilla, S.; Zivkovic, B.; MacKenzie, E. T. *J. Pharmacol. Exp. Ther.* **1988**, *247*, 1222−1232.
(72) Karakas, E.; Furukawa, H. *Science* **2014**, *344*, 992−997.
(73) Lee, C.-H.; Lü, W.; Michel, J. C.; Goehring, A.; J. Du, X. S.; Gouaux, E. *Nature* **2014**, *511*, 191−197.
(74) Weskamp, N.; Hüllermeier, E.; Kuhn, D.; Klebe, G. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2007**, *4*, 310−320.
(75) Ekici, O. D.; Paetzel, M.; Dalbey, R. E. *Protein Sci.* **2008**, *17*, 2023−2037.
(76) Wang, G.; Dunbrack, R. L. *Bioinformatics* **2003**, *19*, 1589−1591.
(77) Rost, B. *Protein Eng.* **1999**, *12*, 85−94.
(78) Ye, Y.; Godzik, A. *Nucleic Acids Res.* **2004**, *32*, 582−5.
(79) Fober, T.; Mernberger, M.; Klebe, G.; Hüllermeier, E. *Efficient Similarity Retrieval for Protein Binding Sites Based on Histogram Comparison*, German Conference on Bioinformatics, Braunschweig, Germany, 2010; pp 51−60.
(80) Wang, R.; Wang, S. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1422−6.
(81) Verdonk, M. L.; Berdini, V.; Hartshorn, M. J.; Mooij, W. T.; Murray, C. W.; Taylor, R. D.; Watson, P. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 793−806.