



# In search of the representative pharmacophore hypotheses of the enzymatic proteome of *Plasmodium falciparum*: a multicomplex-based approach

Anu Manhas<sup>1</sup> · Mohsin Y. Lone<sup>1,2</sup> · Prakash C. Jha<sup>3</sup>

Received: 28 May 2018 / Accepted: 6 October 2018 / Published online: 12 October 2018  
© Springer Nature Switzerland AG 2018

## Abstract

Drug resistance has made malaria an untreatable disease and therefore intensified the need for the development of new drugs and the identification of potential drug targets. In this pursuit, in silico efforts made in the past have not shown significant responses. Therefore, in the present work, the multicomplex-based pharmacophore modeling approach was employed to construct the pharmacophores of the 16 selected *Plasmodium falciparum* (*Pf*) targets. All the constructed hypotheses (153) were screened against a focused dataset made up of experimental actives of the chosen targets (3705 inhibitors). The rationale was to check the affinity of the inhibitors for the off-targets. Subsequently, the constructed hypotheses from each target were pooled based on the feature types and the pooled-hypotheses were then clustered to offer an insight about the pharmacophore similarity. Tanimoto similarity index was also calculated to look for the similarity among the inhibitors belonging to different *Pf* targets. Overall, the work was accomplished to bid healthier perceptive of the pharmacophore-based virtual screening and abet in providing guiding principles for the construction of stringent pharmacophores that can be employed for the screening.

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s11030-018-9885-5>) contains supplementary material, which is available to authorized users.

---

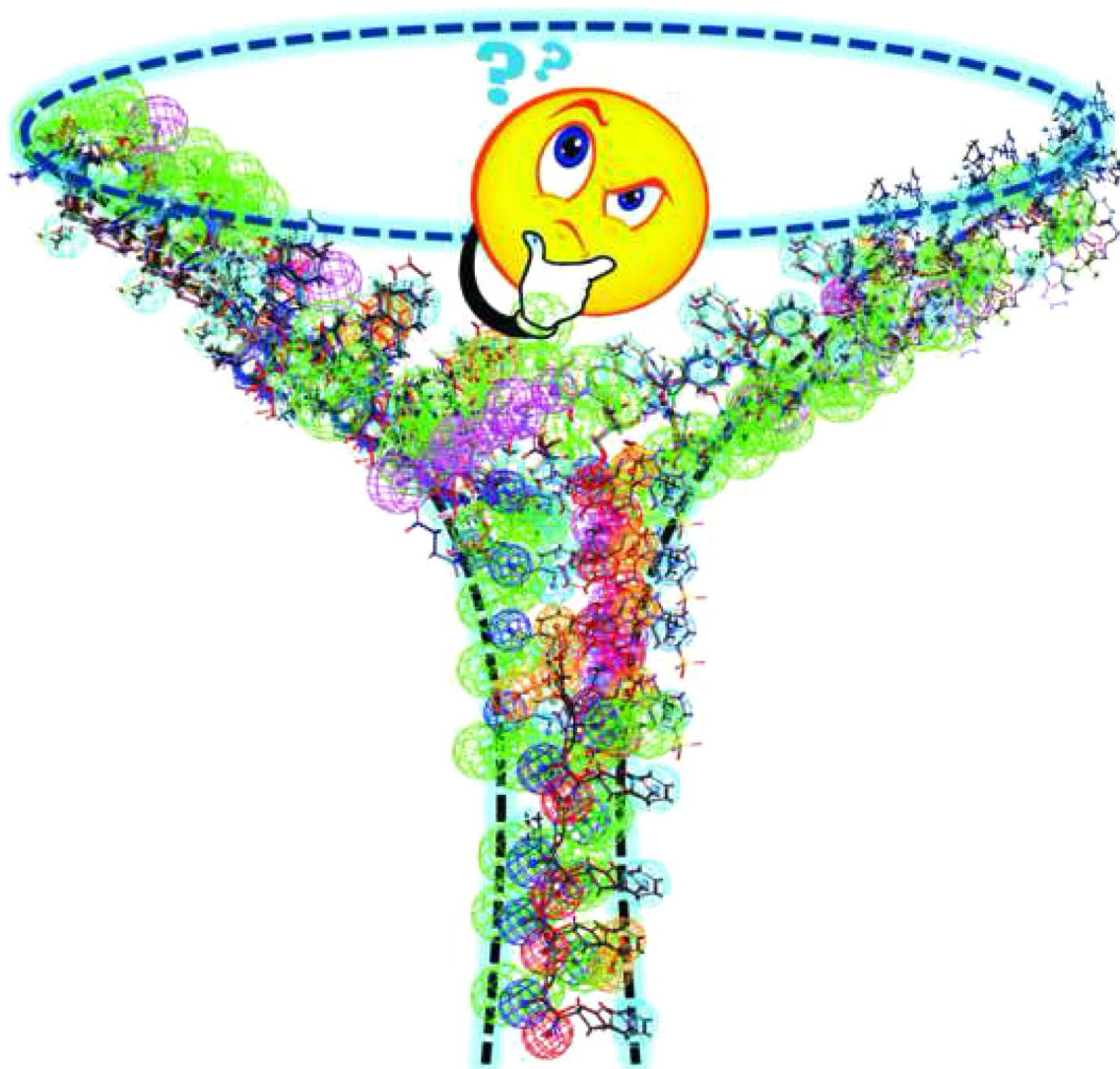
✉ Prakash C. Jha  
prakash.jha@cug.ac.in

<sup>1</sup> School of Chemical Sciences, Central University of Gujarat, Gandhinagar, Gujarat 382030, India

<sup>2</sup> Department of Chemistry, Indian Institute of Technology Gandhinagar, Gandhinagar, Gujarat 382355, India

<sup>3</sup> Centre for Applied Chemistry, Central University of Gujarat, Gandhinagar, Gujarat 382030, India

## Graphical abstract



**Keywords** Multicomplex-based pharmacophore · Enzymatic proteome · Clustering · Tanimoto similarity · Virtual screening

## Introduction

Malaria is a lethal disease caused by *Plasmodium falciparum* (*Pf*), one of the deadly species of the *plasmodium* parasite [1]. According to WHO, around 3.3 billion cases and 438,000 deaths were reported in 2015 [2]. Although there has been a decline in the malaria transmission over the past 15 years (2000–2015) [3], most of the countries face widespread of malaria due to the emergence of drug resistance [4–8]. Owing to this reason, WHO has considered malaria as the first priority tropical disease [9]. Therefore, the scenario has drawn attention of the researchers toward the design of new strategies/drugs to accelerate the process of eradication of malaria.

However, it is a well-known fact that the drug discovery and development is a laborious, time-consuming and an expensive process. Thus, the application of computer-aided molecular design and development of computational methods for lead generation and optimization are of enormous significance to reduce the overall cost and time allied with drug discovery program [10]. In this pursuit, a class of computational techniques known as pharmacophore modeling has been extensively applied [11–13]. This technique has proven to be very effective for lead identification and has been fundamentally divided into structure-based (SB) and ligand-based (LB) approaches. LB methods solely rely on the information extracted from the set of ligands, whereas SB methods

depend on the molecular recognition between the protein–ligand complexes. However, it is well reported in the literature that the LB method suffers from drawbacks as the generation of the model depends on the ligands only [14]. LB method solely depends on the selection of the appropriate training set which affects the performance of the model [15–21]. One way to overcome this drawback is to switch to SB approach which imposes necessary constraints vital for activity and selectivity [14, 22, 23]. In general, SB methods exploit the information of either apo-structure (protein only) or single protein–ligand complex to the hypothesis generation. However, if more than one receptor–ligand complexes are available, then multicomplex-based pharmacophore modeling would be the finest to incorporate all the important interaction patterns simultaneously [24, 25]. Therefore, this strategy was employed to search the representative hypotheses of the selected enzymatic druggable targets of *Pf*.

In the present study, six protein classes of *Pf* viz. Oxidoreductases, Hydrolases, Transferases, Lyases, Isomerases and Ligases were chosen. However, only 16 different enzymes (from five protein classes) were selected on the basis of experimental inhibitory activity. Subsequently, the protein–ligand complexes of 16 groups were subjected to the pharmacophore generation using multicomplex-based approach [24, 26–28]. The generated models were typically validated by using a focussed database made up of experimental actives of the selected targets. Eventually, the generated pharmacophore models from each group were pooled based on the feature types. A model from each pool was selected and then clustered by using Euclidean distance method. The aspiration was to provide an insight about the pharmacophore similarity among the selected enzyme classes and share of features among the inhibitors. Rationale behind the current work was to search for the representative pharmacophores and to provide guiding principle for the design of stringent pharmacophore that can be employed for the virtual screening (VS). We anticipate that the present study (Scheme 1) will advance the knowledge of SB pharmacophore modeling approach as well as provide useful suggestion for their application in VS.

## Materials and methods

### Selection and preparation of protein–ligand complexes

All the protein–ligand complexes were manually selected from the RCSB Protein Data Bank (Supplement Table 1). However, only those complexes were selected which possess the following criteria: (1) the crystallized ligands/inhibitors must possess experimentally determined activity in terms of  $IC_{50}$ ,  $K_i$  or  $K_d$  measures, (2) the protein must belong to the enzymatic protein class, (3) the target must be crystallized

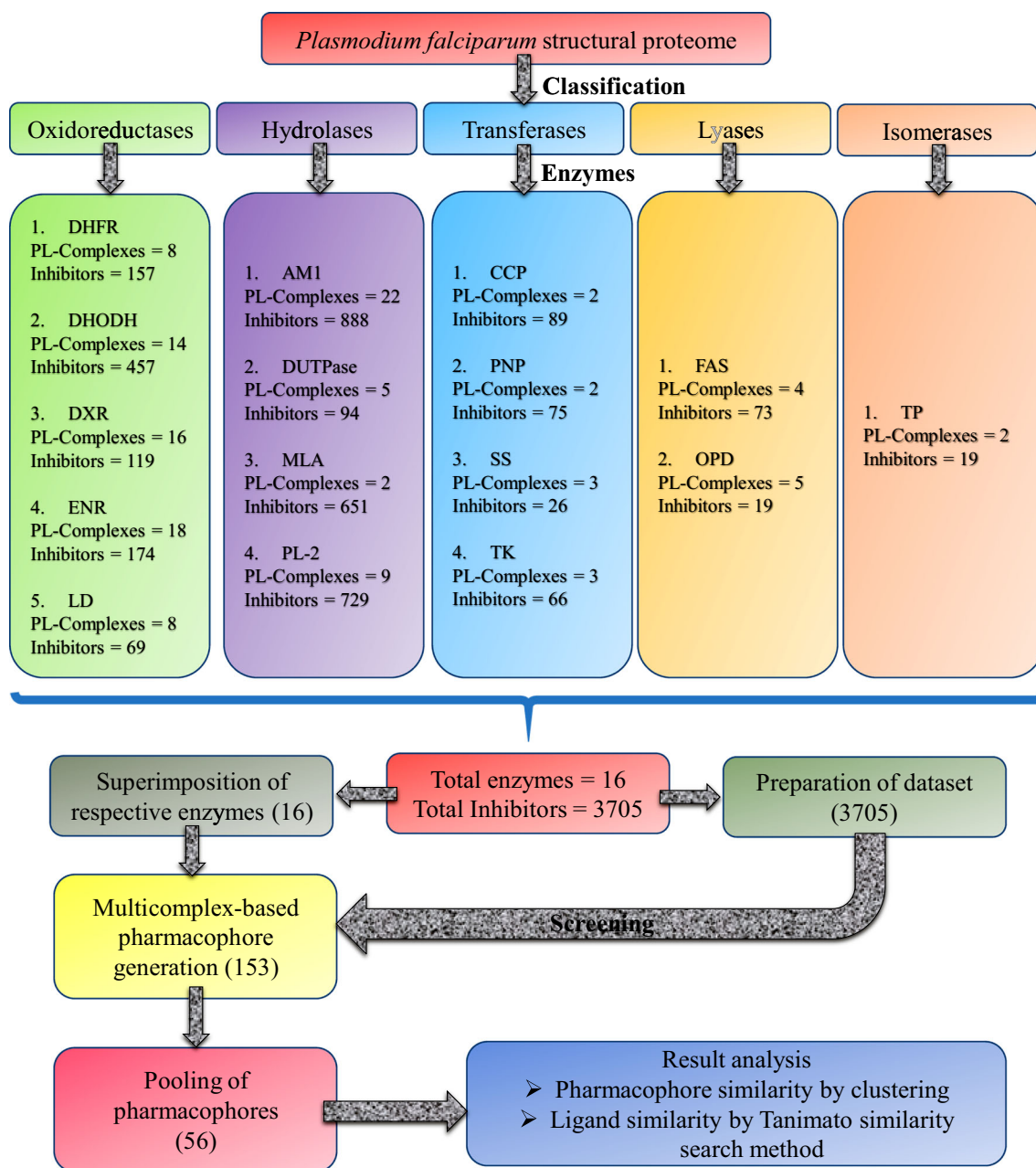
with more than one inhibitors/ligands, (4) the crystallized inhibitors/ligands must belong to the same condensation site and (5) the protein–ligand complex with highest resolution should be taken, if solved at different crystallographic resolutions. This resulted in the selection of five protein classes for the present study (Fig. 1). The selected complexes were retrieved and subsequently grouped on the basis of enzyme commission number. The grouped protein–ligand complexes were then prepared by employing *Protein Preparation Wizard* of Accelrys-Discovery Studio (DS) [29].

### Superimposition and pharmacophore generation

The alignment and superimposition of the prepared protein–ligand complexes from each group were carried out by exploiting *Align and Superimpose Proteins* module of Accelrys-DS [29]. The rationale behind the superimposition was to integrate the essential common interactions in a single coordinate file. The choice of reference for the superimposition of proteins was made on the basis of highest crystallographic resolution (Table 1). Subsequent to the superimposition, all the protein chains were deleted except the reference in order to avoid the repetition of 3D-coordinates of the proteins. The superimposed conformers from each group were then subjected to the pharmacophore generation by using the HipHop algorithm of the *Common Feature Pharmacophore Generation* protocol of Accelrys-DS [29]. Owing to the use of crystal bound conformers, the conformational flexibility of the ligands was disabled prior to pharmacophore construction. In addition, the inter-feature distance was set to 2 Å to consider close chemical features during pharmacophore generation. The pharmacophore features viz. hydrogen-bond acceptor (A), hydrogen-bond donor (D), hydrophobic (H), hydrophobic aliphatic (Z), hydrophobic aromatic (Y), positive ionizable (P), negative ionizable (N) and ring aromatic (R) were requested for the model construction.

### Database preparation and pharmacophore validation

A curated database comprising of 3705 experimental actives of the 16 selected targets was made to test the performance of the generated pharmacophore models (Supplement Table 2). The molecules of the database pertaining to the selected 16 druggable targets were extracted from the literature as well as retrieved from the BindingDB [30] and ChEMBL [31] (release June-2017) databases. Due to the unavailability of the experimental actives of triosephosphate isomerase (TP), human inhibitors were taken into consideration. All the chosen molecules were then collated into a single structure data format (SDF) file which was prepared by employing CHARMM force field [32]-based BEST conformational generation method of *Build 3D Database* protocol



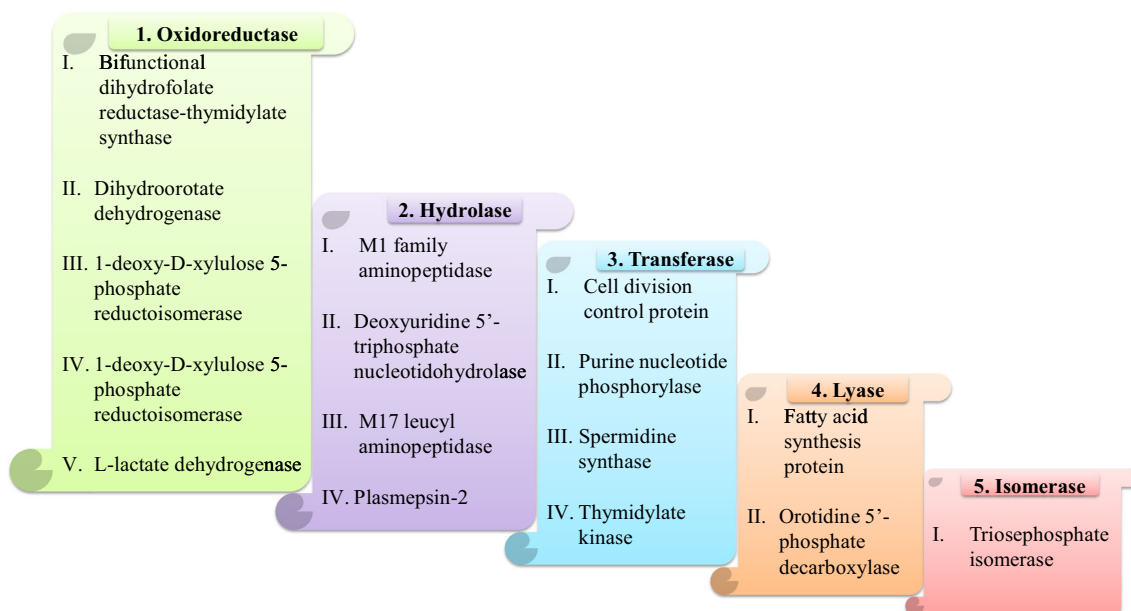
**Scheme 1** Diagrammatic representation of the workflow adopted to conduct the current study (PL-complex is protein–ligand complex)

of Accelrys-DS [29]. This method generates 255 conformers from each molecule within a threshold of 20 kcal/mol of energy above the global minima and eliminates the structural duplicates to maintain the consistency in the dataset. The quality of all the constructed pharmacophores was evaluated by using the prepared database. Each model was mapped at default parameters by employing the *Search 3D Database* module of Accelrys-DS [29]. The statistical parameters viz. Ht (total number of hits retrieved), Ha (total number of actives retrieved), %A (recall of actives), %RA (precision), sensitivity, specificity, area under curve of the receiver operating characteristic plot (AUC-ROC), enrichment factor (EF) and

Güner Henry scores (GH) were calculated. To accomplish these parameters for the hypotheses of a particular target, all the molecules in the database were presumed to be inactive except its own inhibitors.

### Pooling, pharmacophore-clustering and ligand-similarity search

All the constructed hypotheses of the selected *Pf* targets were pooled on the basis of pharmacophore features. The pooling was carried out to avoid the possibility of the repetition of features for a particular target. The pooled-hypotheses from each



**Fig. 1** Pictorial representation of the selected sixteen enzymes of five protein classes of *Plasmodium falciparum* viz. Oxidoreductases, Hydrolases, Transferases, Lyases and Isomerases

target were then subjected to clustering by using the web-interface tool CIMminer [33]. Euclidean distance method and complete linkage algorithm were selected during clustering. The equal width binning algorithm was applied to color the distribution of the resultant clusters. However, to generate feature-based 2D-clustered image map (CIM), the descriptors (D, A, H, Z, Y, P, N and R) of the pooled-hypotheses along with their frequency were compiled in a  $[8 \times 56]$  matrix as an input (Supplement Table 3). The rationale behind the clustering of was to analyze the pharmacophore similarity across the developed models.

On the other hand, the ligand-similarity search was carried out to check the proximity and similarity between the actives and presumed inactives retrieved by a particular hypothesis. To accomplish this, fingerprint-based Tanimoto Similarity Coefficient was calculated by using *Find Similar Molecules by Fingerprints* protocol of Accelrys-DS [29]. The combined clustering and ligand-similarity approach demonstrates the effectiveness to identify and analyze the interesting pattern of the arrangement of the pharmacophores and elucidates the reason accountable for the random distribution of the inhibitors across the generated models.

## Results

### Common feature pharmacophore construction

Multicomplex-based pharmacophore models for sixteen *Pf* enzymes crystallized with different inhibitors from five protein classes were generated by employing the *Common*

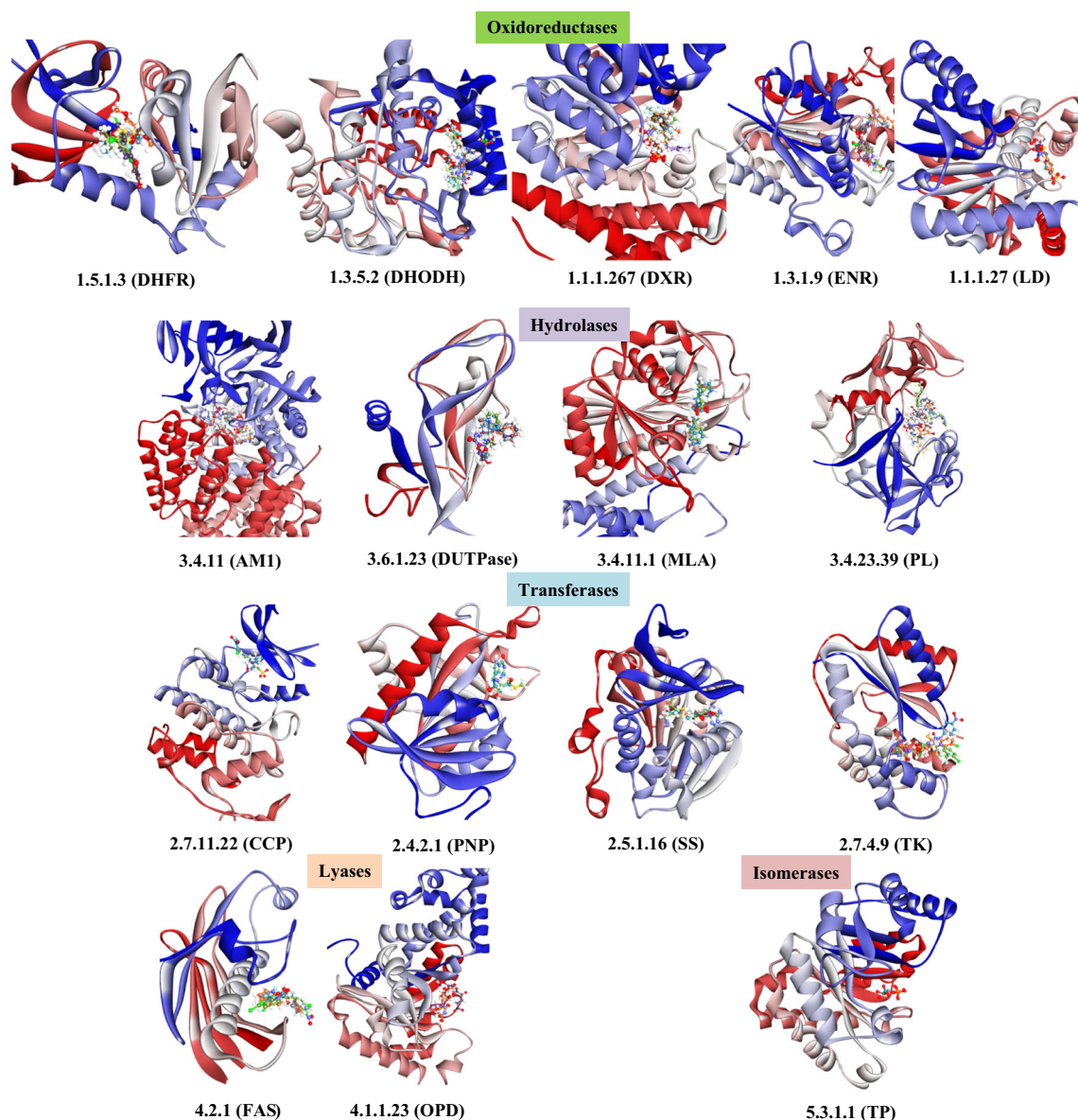
*Feature Pharmacophore Generation* protocol of Accelrys-DS [29]. However, prior to the generation of pharmacophore models, all the protein–ligand complexes of a particular enzyme were superimposed keeping a representative protein as the reference (Table 1). The rationale behind the superimposition was to transform the 3D coordinates of different protein–ligand complexes to a common frame (Fig. 2). Subsequently, the pharmacophores were constructed by exploiting the bioactive conformations of the crystallized ligands of each enzyme. A total of 153 multicomplex-based pharmacophore models were generated (Supplement Table 4 and Supplement Fig. 1) and subsequently subjected to screening by using the prepared focused database. The main objective was to analyze the distribution pattern of the inhibitors across the *Pf* enzymatic proteome. From the screening pattern, it was observed that the molecules have shown affinity for the off-targets (Fig. 3, Supplement Table 5). As stated, it seems nearly impossible for a model to screen its own experimental actives without mapping the large number of presumed inactives.

For instance, hypothesis-3 of L-lactate-dehydrogenase (LD) and all the hypotheses of Orotidine-5'-phosphate-decarboxylase (OPD) retrieved 100% of its actives as well as 88.58% and  $\leq 1.95\%$  of presumed inactives, respectively (Supplement Table 6). However, few models like model-2 of 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR), all models of M17-leucyl aminopeptidase (MLA), cell division control protein (CCP), triosephosphate isomerase (TP) and five models of purine nucleotide phosphorylase (PNP) were unsuccessful in retrieving even a single active molecule from the database. One of the potential ways to overcome this

**Table 1** List of the selected druggable proteins of five enzymatic classes of *Plasmodium falciparum* along with their abbreviation, EC number, PDB IDs and crystallographic resolution

S. No	Druggable targets (EC number)	PDBs [Resolution (Å)]
1	<i>Oxidoreductases</i>	
1a	Bifunctional dihydrofolate reductase-thymidylate synthase (DHFR, 1.5.1.3)	1J3J (2.30), <b>1J3K (2.10)<sup>a</sup></b> , 3DG8 (2.58), 3DGA (2.70), 3JSU (2.70), 3UM8 (2.60), 4DP3 (2.40), 4DPH (2.38)
1b	Dihydroorotate dehydrogenase (DHODH, 1.3.5.2)	5FI8 (2.32), 5TBO (2.15), 5DEL (2.20), 4RX0 (2.25), 4CQ9 (2.72), <b>4CQ8 (1.98)<sup>a</sup></b> , 4CQA (2.82), 4ORM (2.07), 3SFK (2.95), 3O8A (2.30), 3I65 (2.00), 3I68 (2.40), 3I6R (2.50), 1TV5 (2.40)
1c	1-Deoxy-D-xylulose 5-phosphate reductoisomerase (DXR, 1.1.1.267)	<b>5JAZ (1.40)<sup>a</sup></b> , 5JBI (1.70), 5JC1 (1.65), 5JMP (1.70), 5JMW (1.55), 5JNL (1.60), 5JQ0 (1.80), 4Y67 (1.60), 4Y6P (1.90), 4Y6R (1.90), 4Y6S (2.10), 3WQQ (2.25), 3WQR (1.97), 3WQS (2.35), 4KP7 (2.00), 4GAE (2.30)
1d	Enoyl-acyl carrier reductase (ENR, 1.3.1.9)	4IGE (2.15), 4IGF (2.30), 2OP0 (2.80), 2OP1 (2.60), 2FOI (2.50), 2OL4 (2.26), 2OOS (2.10), 1ZW1 (2.90), 1NHW (2.35), 1NNU (2.50), 3AM5 (2.05), 3LSY (2.85), <b>3LT0 (1.96)<sup>a</sup></b> , 3LT1 (2.20), 3LT2 (2.50), 3LT4 (2.25), 1ZXB (2.68), 1ZXL (3.00), 1ZSN (2.99)
1e	L-lactate dehydrogenase (LD, 1.1.1.27)	<b>4PLZ (1.05)<sup>a</sup></b> , 1U40 (1.70), 1U4S (2.00), 1U5A (1.80), 1T24 (1.70), 1T25 (1.90), 1T26 (1.80), 1CET (2.05)
2	<i>Hydrolases</i>	
2a	M1 family aminopeptidase (AM1, 3.4.11)	4ZW3 (1.80), 4ZW5 (1.80), 4ZW6 (1.90), 4ZW7 (1.95), 4ZW8 (2.00), 4ZX3 (2.00), 4ZX4 (1.90), 4ZX5 (1.95), 4ZX6 (2.05), 4R5T (1.98), 4R5V (2.10), 4R5X (1.85), 3T8V (1.80), 4K5L (1.91), 4K5M (1.75), 4K5N (1.91), 4K5O (1.90), 4K5P (1.85), 3Q43 (1.80), 3Q44 (1.80), <b>3EBH (1.65)<sup>a</sup></b> , 3EBI (2.00)
2b	Deoxyuridine 5'-triphosphate nucleotidohydrolase (DUTPase, 3.6.1.23)	3T60 (2.40), 3T6Y (2.60), <b>3T70 (1.80)<sup>a</sup></b> , 2Y8C (2.10), 1VYQ (2.40)
2c	M17 leucyl aminopeptidase (MLA, 3.4.11.1)	<b>3KR4 (2.00)<sup>a</sup></b> , 3KR5 (2.56)
2d	Plasmepsin-2 (PL-2, 3.4.23.39)	4Z22 (2.62), 4CKU (1.85), 1W6H (2.24), <b>2BJU (1.56)<sup>a</sup></b> , 1LF2 (1.80), 1LF3 (2.70), 1LEE (1.90), 4YA8 (3.30), 4Y6M (2.27)
3	<i>Transferases</i>	
3a	Cell division control protein (CCP, 2.7.11.22)	<b>1V0O (1.90)<sup>a</sup></b> , 1V0P (2.00)
3b	Purine nucleotide phosphorylase (PNP, 2.4.2.1)	1NW4 (2.20), <b>1Q1G (2.02)<sup>a</sup></b>
3c	Spermidine synthase (SS, 2.5.1.16)	4CXM (1.75), 2PT9 (2.20), <b>2I7C (1.71)<sup>a</sup></b>
3d	Thymidylate kinase (TK, 2.7.4.9)	2YOF (1.82), <b>2YOG (1.50)<sup>a</sup></b> , 2YOH (1.60)
4	<i>Lyases</i>	
4a	Fatty acid synthesis protein (FAS, 4.2.1)	3AZ8 (3.10), 3AZ9 (2.75), 3AZA (2.70), <b>3AZB (2.60)<sup>a</sup></b>
4b	Orotidine 5'-phosphate decarboxylase (OPD, 4.1.1.23)	3VI2 (2.10), 3S9Y (1.70), <b>3N3M (1.47)<sup>a</sup></b> , 2Q8Z (1.80), 3BAR (1.90)
5	<i>Isomerases</i>	
5a	Triosephosphate isomerase (TP, 5.3.1.1)	2VFI (2.25), <b>1LYX (1.90)<sup>a</sup></b>

<sup>a</sup>Reference selected for superimposition



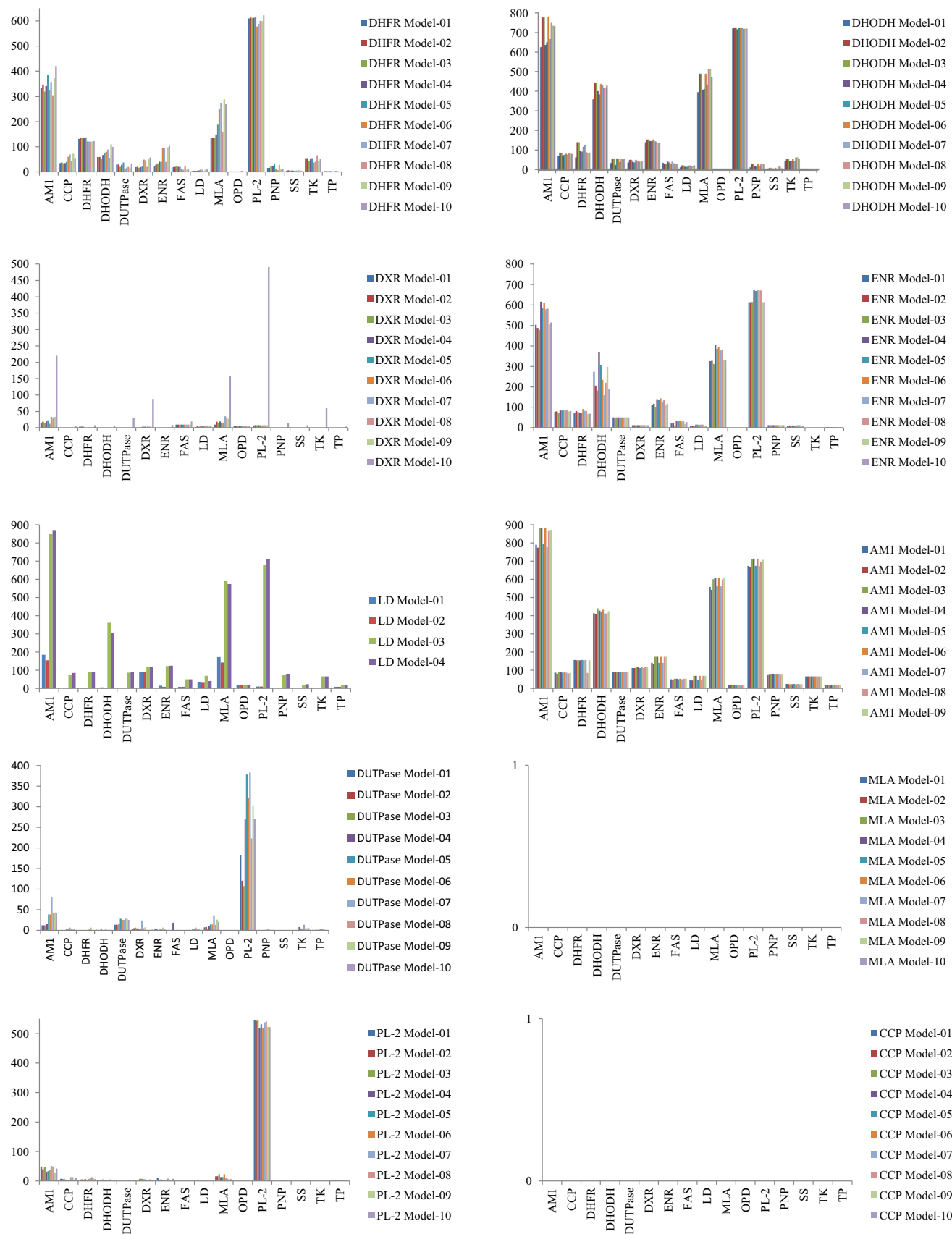
**Fig. 2** Superimposed protein–ligand complexes of selected druggable proteins of five enzymatic classes of *Plasmodium falciparum*

problem is to impose necessary constraints to the generated stringent-pharmacophore models. Keeping this in view, the current study was accomplished by exploiting *Pf* enzymatic proteome and thus expands the application domain of the pharmacophore modeling.

### Validation metrics

The main objective of the pharmacophore-based virtual screening is to increase the probability of retrieving such molecules from the database which are likely to be active in the experimental findings. Accordingly, virtual screening (VS) is considered as a promising technique for the filtering of large set of molecules in the database. Thus, selection

of the model for conducting the VS becomes an essential step. However, to evaluate the quality of the hypothesis, various quality metrics such as EF, %A, %RA, specificity, sensitivity, GH and AUC-ROC were used. To obtain these parameters, the common practice is to use a validation set comprising of experimentally confirmed inhibitors (actives) and experimentally reported non-inhibitors (inactives) of the particular target. In contrast, the present work exploited a focussed dataset made up of the experimental actives of the 16 selected enzymes. The aspiration was not to statistically validate the pharmacophores rather to check the real-world distribution of the inhibitors and their affinity for the off-targets. Thus, all the pharmacophores (153) were initially assessed by employing a test set comprising of the actives



**Fig. 3** Graphical representation of the number of inhibitors mapped by a particular pharmacophore model from the focused dataset



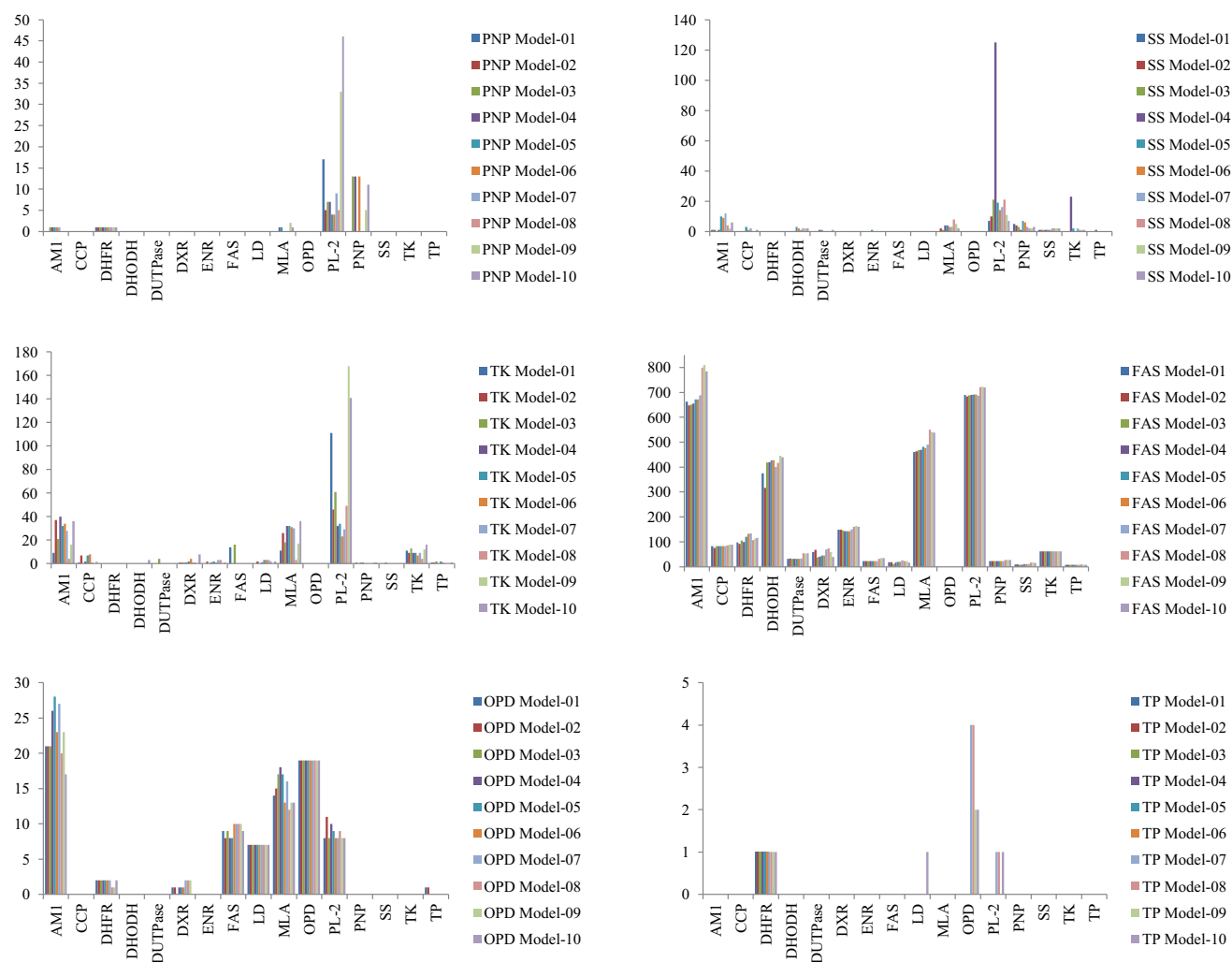


Fig. 3 continued

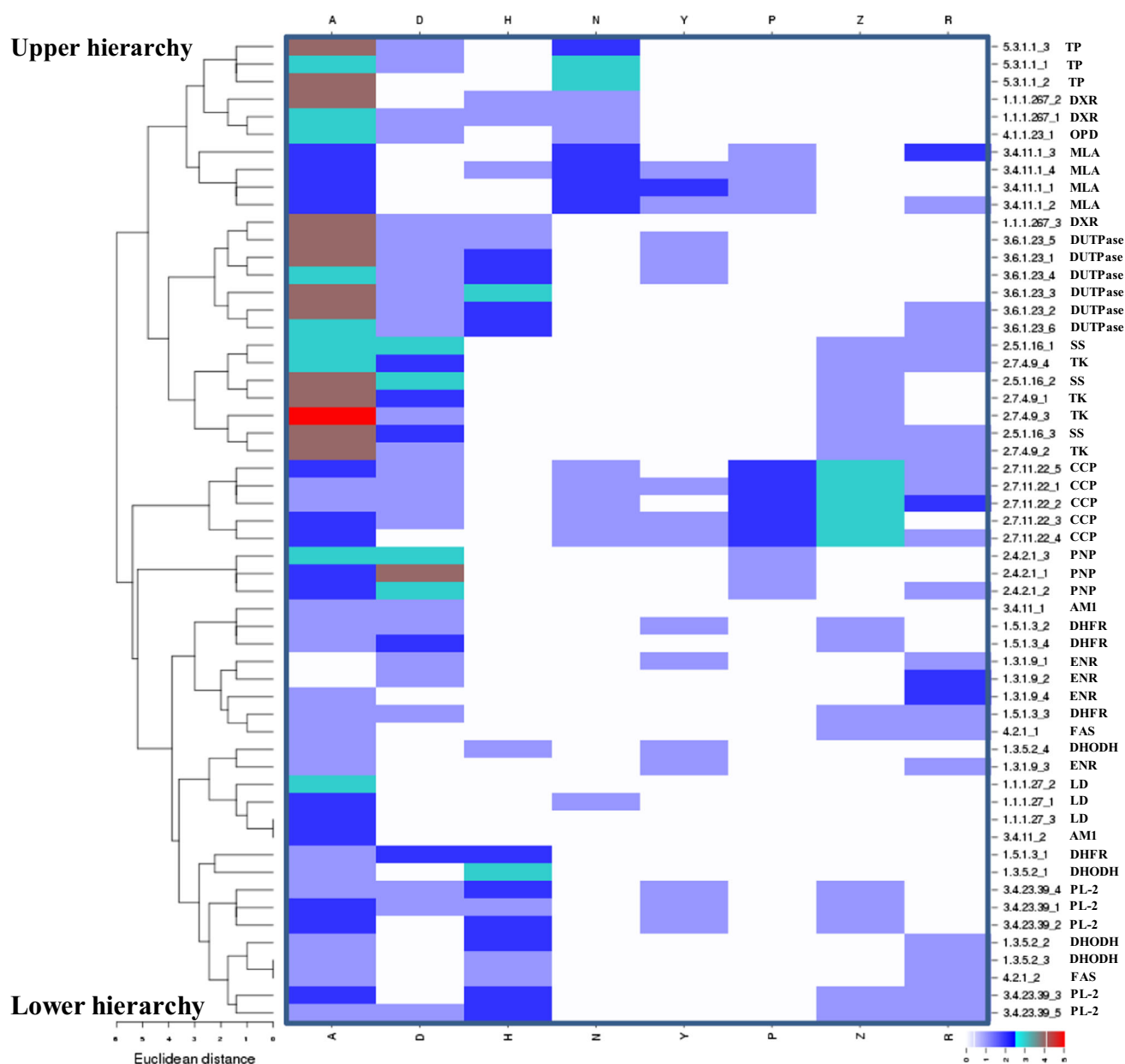
and presumed inactive. This method was used to estimate the sensitivity, specificity and AUC-ROC plots of all the hypotheses. In general, sensitivity and specificity indicates the ability of the model to identify the actives and to exclude the inactives from a dataset, respectively, whereas AUC-ROC predict the accuracy of the model to pick actives prior to presumed inactive. From the Supplement Table 6, it is obvious that a wide range of values were obtained for the above-mentioned parameters. Except the hypotheses obtained for enzymes Plasmepsin-2 (PL-2) and OPD, all have displayed discrepancies in terms of sensitivity, specificity and AUC.

In addition, the EF and its descriptors (%A and %RA) were also calculated. Generally, EF quantifies the recognition of the actives explicitly compared to the presumed inactive and its higher value corresponds to the better performance and reliability of the model. It is evident from the Supplement Table 6 that the diverse range (0–48.85) of EF values was shown by the hypotheses of the selected targets. However, a significantly good EF values were obtained for all

the hypotheses of PL-2 (4.16–4.54) and OPD (40.26–48.85) enzymes. This behavior was in one-to-one correlation with the sensitivity, specificity and AUC values of the hypotheses of PL-2 and OPD. In general, the criteria for the ideal model are to display sensitivity and specificity  $\approx 1$ , high EF and AUC value with a steep slope of ROC curve. It is imperative to mention that the models obtained from the OPD fits well within the ideal model criteria and thus can be used for the VS (Supplement Table 6 and Supplement Fig. 2). However, the rest of hypotheses from other targets have shown affinity for the diverse inhibitors thus necessitates the use of stringent-pharmacophore models. To accomplish this, pharmacophore-clustering and Tanimoto-based ligand-similarity studies were conducted.

### Cluster and similarity analysis

The similarity between the pharmacophore models of the selected protein targets was provided by carrying out phar-



**Fig. 4** Dendrogram representing the feature-based 2D-clustered image map of pooled-pharmacophores based on the composition of the features

macophore clustering. The key objective was to study the distinct patterns and their relation with the specificity of the developed models. However, instead of comparing 3D-coordinates of the feature types, the composition of features was chosen. The genesis of this postulate stems in the literature where 3D-pharmacophore-based clustering was performed to understand the intermolecular interactions [34].

Starting from the lower hierarchy of the dendrogram (Fig. 4), the pooled-hypotheses of PL (5 pools) were found to form two clusters, comprising of two and three hypotheses, respectively. None of the hypotheses were observed to cluster with the hypotheses of other targets, thus displaying

the distinct pattern of feature types. Essentially, all the clustered hypotheses comprise of six features; however, the lower hierarchical cluster differs with respect to one feature, while the higher hierarchical cluster differs by two features. It is imperative to mention that 71.33–75.08% of PL-2 inhibitors were retrieved from the database via clustered hypotheses (Table 2). This clearly indicates the selectivity of these models toward its own inhibitors. Despite the significant specificity, these six-feature pharmacophore models were found to extract the small number of the inhibitor molecules from the off-targets such as aminopeptidase (AM1), dihydrofolate reductase (DHFR), dihydroorotate dehydrogenase

**Table 2** List of the number of inhibitors retrieved by the pooled-hypotheses from the prepared focussed dataset

Cluster	Features	AMI	CCP	DHFR	DHODH	DUTPase	DXR	ENR	FAS	LD	MLA	OPD	PL	PNP	SS	TK	TP
<i>Oxidoreductases</i>																	
Bifunctional dihydrofolate reductase-thymidylate synthase (DHFR, 1.5.1.3)																	
1	HHDDA	332	35	132	60	30	19	23	19	3	134	0	610	16	4	55	3
2	YZDA	324	61	121	80	14	49	94	12	9	250	0	578	14	4	38	2
3	RZDA	358	70	121	89	17	47	95	9	9	273	0	588	8	6	41	3
4	ZDDA	305	42	120	56	20	21	40	22	3	160	4	600	29	7	66	5
Dihydroorotate dehydrogenase (DHODH, 1.3.5.2)																	
1	HHHA	623	68	61	357	32	35	138	33	11	394	0	719	11	4	43	3
2	RHHA	774	84	138	441	54	48	152	27	20	487	0	723	26	7	52	4
3	RHA	747	82	88	418	52	40	141	31	17	511	0	716	26	14	61	3
4	YHA	730	78	85	427	52	41	136	29	20	469	0	717	26	6	52	4
1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR, 1.1.1.267)																	
1	NHDDAA	15	0	3	0	0	1	0	10	2	10	5	6	0	0	0	0
2	NHAAAA	12	0	1	0	0	2	0	10	5	16	6	7	0	0	0	0
3	HDAAAA	220	6	8	6	30	88	9	19	7	159	6	491	14	8	60	3
Enoyl-acyl carrier reductase (ENR, 1.3.1.9)																	
1	RYD	504	79	72	274	50	13	111	21	8	325	0	613	13	10	0	2
2	RRD	488	80	81	206	46	12	117	23	8	327	0	613	13	11	0	2
3	RYA	617	84	74	371	50	13	140	33	16	406	0	675	13	11	0	3
4	RRA	580	84	83	159	50	13	124	31	16	378	0	675	13	12	0	3
L-lactate dehydrogenase (LD, 1.1.1.27)																	
1	NAA	185	0	2	6	0	91	16	10	34	172	19	11	0	0	0	9
2	AAA	849	73	89	361	87	119	123	51	69	590	19	678	76	20	66	20
3	AA	871	86	92	308	91	119	125	51	41	575	19	713	80	24	66	18
<i>Hydrolases</i>																	
M1 family aminopeptidase (AMI, 3.4.11)																	
1	DA	790	87	157	414	91	113	141	51	49	559	19	674	77	25	66	18
2	AA	881	88	155	441	91	119	174	54	69	601	19	713	80	24	66	20
Deoxyuridine 5'-triphosphate nucleotidohydrolase (DUTPase, 3.6.1.23)																	
1	YHDDAAA	12	0	0	1	13	2	0	0	0	7	0	183	0	0	0	1
2	RHDDAAA	12	0	0	1	13	5	1	0	0	8	0	120	0	0	0	0
3	HHDDAAA	16	0	0	2	16	4	2	18	0	10	0	269	0	0	8	1
4	YHDDAAA	38	3	0	0	28	4	3	0	3	14	0	378	1	0	4	1
5	YHDDAAA	79	6	1	2	25	24	2	1	6	36	0	383	3	0	13	2
6	RHDDAAA	41	2	3	0	28	5	3	0	2	13	0	223	1	0	4	2
M17 leucyl aminopeptidase (MLA, 3.4.11.1)																	
1	PNNYAA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	RPNYAA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 2 continued

Cluster	Features	AMI	CCP	DHFR	DHODH	DUTPase	DXR	ENR	FAS	LD	MLA	OPD	PL	PNP	SS	TK	TP
3	RRNNAA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	PNNYHAA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Plasmeprin-2 (PL-2, 3.4.23.39)																	
1	YZHDA	49	8	5	2	0	2	12	0	0	16	0	547	0	0	0	2
2	YZHAA	40	7	6	2	0	8	4	0	1	16	0	543	0	0	0	2
3	RZHAA	31	5	7	3	0	6	4	0	1	13	0	520	0	0	0	2
4	YZHDA	51	13	11	6	3	5	9	0	0	10	0	538	1	0	0	2
5	RZHDA	27	6	9	0	0	3	3	0	1	5	0	523	0	0	0	2
<i>Transferases</i>																	
Cell division control protein (CCP, 2.7.11.22)																	
1	RPPNYZZDA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	RRPPNZZDA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	PNNYZZDAA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	RPPNYZZAA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	RPPNZZDAA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Purine nucleotide phosphorylase (PNP, 2.4.2.1)																	
1	PDDDDAA	0	0	1	0	0	0	0	0	0	0	0	17	0	0	0	0
2	RPDDDAA	1	0	1	0	0	0	0	0	0	0	0	7	13	0	0	0
3	PDDDDAA	0	0	1	0	0	0	0	0	0	2	0	33	5	0	0	0
Spermidine synthase (SS, 2.5.1.16)																	
1	RZDDDAA	1	0	0	0	0	0	0	0	0	0	0	7	5	1	0	0
2	ZDDDDAA	1	0	0	0	1	0	0	0	0	4	0	125	1	1	23	1
3	RZDDAAA	10	3	0	3	1	0	1	0	0	4	0	19	7	1	2	0
Thymidylate kinase (TK, 2.7.4.9)																	
1	ZDDAAA	9	1	0	0	0	1	0	14	0	11	0	111	1	0	11	1
2	RZDAAA	37	7	0	0	0	1	2	0	2	26	0	46	0	0	9	1
3	ZDAAAA	21	0	0	0	4	1	0	16	0	18	0	61	1	1	13	2
4	RZDDAAA	16	2	0	0	0	0	0	0	0	17	0	168	1	0	12	0
<i>Lyases</i>																	
Fatty acid synthesis protein (FAS, 4.2.1)																	
1	RZA	662	82	97	375	31	58	148	23	18	460	0	689	22	9	61	7
2	RHA	799	85	106	417	54	73	161	32	23	550	0	721	27	17	61	9
Orotidine 5'-phosphate decarboxylase (OPD, 4.1.1.23)																	
1	NDA	21	0	2	0	0	1	0	9	7	14	19	8	0	0	0	1
<i>Isomerases</i>																	
Triosephosphate isomerase (TP, 5.3.1.1)																	
1	NNNDAAA	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	NNNAAA	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
3	NNDAAA	0	0	1	0	0	0	0	0	0	0	4	1	0	0	0	0

(DHODH), enoyl-acyl carrier reductase (ENR), DXR, CCP, LD, MLA and TP (Table 2). This concern of non-specificity can be resolved by incorporating either P or N to the six-feature hypothesis. The next lower hierarchical clustered target seen in the cladogram was fatty acid synthesis protein (FAS). All the hypotheses pertaining to this target were pooled into two pools, and the hypothesis from each pool was observed to cluster with DHODH and DHFR, respectively. Owing to the simple pharmacophoric requirements (3-feature), these clustered hypotheses were able to retrieve the inhibitors from all the targets except OPD (Table 2). The demonstration of pharmacophore specificity/sensitivity using 3-feature hypotheses may not be the sufficient indicator to recognize the molecules from pharmacophore-based VS [25]. Traversing forward from FAS, the four pooled-hypotheses of DHODH were found to assemble into three clusters with FAS, DHFR and ENR. The hypotheses clustered with FAS consist of 3 (RHA) and 4 features (RHHA) of DHODH differing with respect to H feature. However, the pooled-hypotheses clustered with DHFR and ENR comprise of AHHH and AHY features, respectively. All the clustered hypotheses were successful in retrieving 78.12–96.50% of its own inhibitors along with the significant contribution from all chosen targets except OPD (Table 2). The lack of sensitivity can be attributed to the simple pharmacophoric requirements. From the cladogram (Fig. 4), it is expected that the removal of A and addition of R to the clustered hypotheses may increase both the specificity and sensitivity. Recently, a study has shown the importance of such features in enhancing the specificity of the pharmacophores of *Pf*DHODH [24]. Similarly, the pooled-hypotheses of DHFR were found to cluster with DHODH, FAS, ENR and AM1 and thereby formed four clusters (Fig. 4). The retrieved 5- and 4-feature hypotheses showed difference of H feature. Akin to DHODH, the high specificity (76.43–84.08%) and low sensitivity were prevalent in all the pooled-hypotheses of DHFR. However, it is clear from the dendrogram that the presence of N/P in conjunction with R may enhance the sensitivity of the clustered-pharmacophore models.

For the sake of brevity, the discussion pertaining to non-meaningful pooled-hypotheses was not made. These include all the pooled-hypotheses of AM1, CCP, spermidine synthase (SS), pooled-hypothesis-3 of LD and 1, 2, 3 of deoxyuridine-5'-triphosphate nucleotidohydrolase (DUTPase). For such targets, it is imperative to prioritize the limited number of chemical features (typically 3–7) to construct a practical hypothesis for the VS experiments [14]. All the meaningful hypotheses of LD were seen to be grouped in the same cluster and differed by one feature type. It is obvious from the results that the N feature has shown dramatic effect on both sensitivity and specificity of the models (Table 2). However, due to simple pharmacophoric entailments, the sensitivity was very low as compared to specificity. We expect the inser-

tion of H and donor features to the existing features may balance the sensitivity and specificity of the models for this target. The adjoining clusters target ENR dwells one, 4 features and three, 3 feature pooled-hypotheses were observed to be grouped into three different clusters. The two lower hierarchical pooled-hypotheses were clustered with DHODH, FAS and DHFR, whereas the upper hierarchical hypotheses were not found to be clustered with any of the hypotheses. Akin to the above-mentioned targets, the models of ENR were not able to discriminate between the actives and inactives thus resulted in insignificant specificity and sensitivity (Table 2). However, clustering analysis suggests that the addition of H and N to the obtained pharmacophores may enhance the sensitivity of the ENR models.

Likewise, the clustering of pharmacophores displayed a single cluster of the pooled-hypotheses of PNP. Each hypothesis comprises of 7-feature types. Despite the significant specificity, these models have shown insignificant sensitivity ranging from 0 to 17.33%. The obvious reason seems to be the complexity of 7-feature hypotheses and the removal of either A, D or R feature may increase the sensitivity. In contrast, the 7-feature pooled-hypotheses of TK were found to cluster into three small clusters with SS only. The sheer dominance of A and D features has made these hypotheses highly specific. Correspondingly, these features were also accountable for low sensitivity ranging from 13.64 to 19.70%. The removal of additional D/A features may lead to augment the sensitivity of these models.

All the significant pooled-hypotheses of DUTPase comprise of 7 features and were found to form three clusters. The first two clusters from the lower hierarchy solely consist of the significant and non-significant pharmacophores of the same target, whereas the last cluster comprises of one significant hypothesis each of DUTPase and DXR. The significant hypotheses have notably excluded the inactives and actives (Table 2). The obtained clustering pattern suggests the omission of excessive D/A features in order to increase the sensitivity of the significant pharmacophores. Adjacent to this cluster, the six-feature pharmacophores of DXR were found to form two clusters, one with the aforementioned target and the other with OPD. The simple pharmacophore requirement of the lower hierarchical clustered hypothesis was accountable for insignificant sensitivity and specificity. However, the presence of N feature has drastically increased the specificity of these models (Table 2). It is therefore expected that the insertion of R feature instead of A may balance the sensitivity and specificity of these models.

Similarly, all the pooled-hypotheses of MLA were found to form a single cluster with each other only. Owing to the presence of a group of complicated features, none of the clustered hypotheses were able to extract any molecule from the focused database. Thus, we expect the removal of single N and P features from the clustered hypotheses may

**Table 3** Table representing the percentage similarity between the mapped inhibitors of the same and different enzymatic proteins from the generated pooled-hypotheses of *Plasmodium falciparum* proteome

Pharmacophore cluster	Mapped molecules	Mapped form same inhibitors	Mapped from other	Similar ligands	% Similarity <sup>a</sup> search
<i>Oxidoreductases</i>					
Bifunctional dihydrofolate reductase-thymidylate synthase (DHFR, 1.5.1.3)					
1	1475	132	1343	1217	90.618
2	1650	121	1529	1436	93.918
3	1734	121	1613	1524	94.482
4	1500	120	1380	1230	89.130
Dihydroorotate dehydrogenase (DHODH, 1.3.5.2)					
1	2532	357	2175	2164	99.494
2	3037	441	2596	2583	99.499
3	2947	418	2529	2507	99.130
4	2872	427	2445	2436	99.632
1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR, 1.1.1.267)					
1	50	1	49	28	57.143
2	59	2	57	34	59.649
3	1134	88	1046	1021	97.610
Enoyl-acyl carrier reductase (ENR, 1.3.1.9)					
1	2095	111	1984	1969	99.244
2	2027	117	1910	1866	97.696
3	2506	140	2366	2359	99.704
4	2221	124	2097	2058	98.140
L-lactate dehydrogenase (LD, 1.1.1.27)					
1	555	34	521	434	83.301
2	3290	69	3221	3029	94.039
3	3279	41	3238	2866	88.511
<i>Hydrolases</i>					
M1 family aminopeptidase (AM1, 3.4.11)					
1	3331	790	2541	2484	97.757
2	3595	881	2714	2652	97.716
Deoxyuridine 5'-triphosphate nucleotidohydrolase (DUTPase, 3.6.1.23)					
1	219	13	206	203	98.544
2	160	13	147	144	97.959
3	346	16	330	324	98.181
4	483	28	455	452	99.341
5	582	25	557	540	96.948
6	327	28	299	295	98.662
M17 leucyl aminopeptidase (MLA, 3.4.11.1)					
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
Plasmepsin-2 (PL, 3.4.23.39)					
1	640	547	93	92	98.925
2	625	543	82	81	98.780

**Table 3** continued

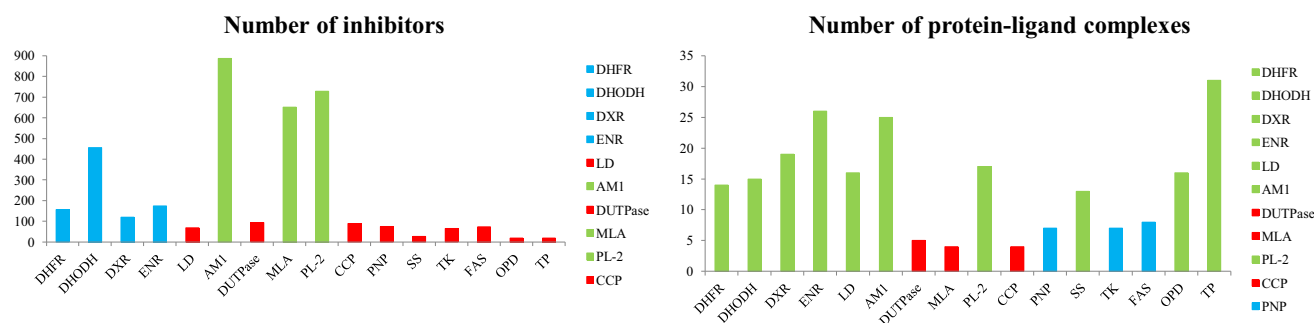
Pharmacophore cluster	Mapped molecules	Mapped from same inhibitors	Mapped from other	Similar ligands	% Similarity <sup>a</sup> search
3	586	520	66	65	98.485
4	645	538	107	104	97.196
5	574	523	51	51	100.000
<i>Transferases</i>					
Cell division control protein (CCP, 2.7.11.22)					
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0
Purine nucleotide phosphorylase (PNP, 2.4.2.1)					
1	18	0	18	0	0
2	22	13	9	0	0
3	41	5	36	0	0
Spermidine synthase (SS, 2.5.1.16)					
1	14	1	13	2	15.385
2	168	1	167	1	0.599
3	51	1	50	2	4.000
Thymidylate kinase (TK, 2.7.4.9)					
1	160	11	149	90	60.403
2	131	9	122	54	44.262
3	138	13	125	71	56.800
4	216	12	204	102	50.000
<i>Lyases</i>					
Fatty acid synthesis protein (FAS, 4.2.1)					
1	2742	23	2719	2126	78.191
2	3135	32	3103	2489	80.213
Orotidine 5'-phosphate decarboxylase (OPD, 4.1.1.23)					
1	82	19	63	6	01.587
<i>Isomerases</i>					
Triosephosphate isomerase (TP, 5.3.1.1)					
1	1	0	1	0	0
2	1	0	1	0	0
3	6	0	6	0	0

$$^a\% \text{Similarity search} = \frac{\text{Similar ligands}}{\text{Mapped from other}}$$

enhance the sensitivity of the models. However, the increase in sensitivity may be at the cost of decrease in specificity of these pharmacophores. Traversing forward, a single pooled-hypothesis obtained for OPD was found to cluster with DXR. The hypothesis consists of NDAAA features and showed the momentous discrimination between the experimental actives and inactives. Therefore, this hypothesis can be employed for the VS experiments. The last hierarchical cluster comprises of three 7-feature pooled-hypotheses of TP, and all the hypotheses were not able to retrieve a single active molecule from the database. The obvious reason seems to be the pres-

ence of excessive N features. Thus, we expect the omission of excess N may enhance the sensitivity of the models.

Beyond the pharmacophore similarity, the other reason that may be accountable for low sensitivity and specificity of the pooled-hypotheses is the similarity among the inhibitors. It is obvious from the results (Table 3) that the fingerprint similarity between the actives and inactives recognized by the pooled-hypotheses ranges from 0 to 100%. Most of the pooled-hypotheses of the chosen targets showed >50% similarity between the mapped actives and inactives. This necessitates the identification of structurally diverse candi-



**Fig. 5** Graphical representation of the number of inhibitors and number of crystallized protein–ligand complexes of the respective enzymes of *Plasmodium falciparum*. Green, blue and red colors represent highly, moderately and least explored targets, respectively

dates for the effective inhibition. However, the inhibitors of some targets like MLA, CCP, PNP, SS, OPD and TP have not shown the considerable similarity (0–15%) between the mapped actives and inactives. The primary reason seems to be the inability of the pooled-hypotheses to retrieve its own inhibitors which in turn can be related to the intricacy in the pharmacophoric features of the generated models. On the other hand, the pooled-hypotheses of PNP and SS were able to retrieve extremely fewer corresponding inhibitors and therefore displayed very low similarity between the actives and inactives.

## Discussion

### Distribution of *Pf* structural data

The reliability of a pharmacophore model on the basis of statistical parameters can only be achieved, if there is a balance in the pharmacophoric requirements, adequate number of inhibitors in the dataset and satisfactory number of co-crystallized protein targets for the model construction [25]. Therefore, the selected *Pf* targets were classified into three groups viz. highly explored, moderately explore and least explored, based on the number of inhibitors and the crystallized protein–ligand complexes. The threshold for the categorization was manually chosen based on the overall distribution.

### Classification on the basis of number of inhibitors

All those targets for which more than 500 inhibitors were reported in the literature were categorized into highly explored targets (AM1, MLA and PL-2). Similarly, the targets for which 100–500 and less than 100 inhibitors were reported were categorized into moderately explored (DHFR, DHODH, DXR and ENR) and least explored (most of the targets), respectively (Fig. 5). For the last two categories

attention has to be paid toward the synthesis of diverse inhibitors.

### Classification on the basis of number of crystallized protein–ligand complexes

The targets which possess more than 10 crystallized protein–ligand complexes were classified as highly explored, whereas the targets crystallized with 5–9 and less than 5 were classified into moderately explored and least explored, respectively (Fig. 5). Majority of the targets have been crystallized with more than 10 different inhibitors thus belong to the highly explored category. On the other hand the targets viz. PNP, TK, FAS and DUTPase, MLA, CCP, fall within the domain of moderately explored and least explored category, respectively.

Overall, the targets viz. PL-2 and AM1 qualified both in terms of number of inhibitors and the crystallized protein–ligand complexes. However, owing to the simple pharmacophoric requirement AM1 was not able to discriminate between actives and inactives. On the other hand the systems of concern include DUTPase and CCP due to their protein–ligand and inhibitor deficiency. The resolution of these deficiencies is a challenge and utmost concern for the generation of multicomplex-based pharmacophore models in order to reflect quantitative structure–activity relationship.

## Conclusions

In summary, the multicomplex-based pharmacophore modeling approach was exercised to construct the pharmacophore models of the 16 selected *Pf* targets. A total of 158 hypotheses were generated and subsequently screened against a focussed database made up of experimental actives and inactives. It was observed that most of the inhibitors have shown affinity for the off-targets. Therefore, various statistical parameters were calculated and correlated with the robustness of the generated models. Subsequently, all the



generated models were pooled and then clustered to analyze the pharmacophore similarity across the selected *Pf* targets. The essential features accountable for the specificity were prioritized, and the rationale behind the non-specificity was highlighted. Both pharmacophore and ligand similarities were found to be accountable for the present distribution of actives and inactives with the dominance of later. Based on pharmacophore clustering and ligand similarities, the solutions were offered to reduce the off-target affinities. Additionally, the promising targets and the valid pharmacophores that can be employed for the virtual screening were highlighted. Overall, the study emphasized the need for the construction of stringent pharmacophore models and the synthesis of structurally diverse inhibitor molecules. Despite the advances in recent past for the combat of malaria, still limited wet lab and molecular modeling efforts have been endeavored for the development of efficient inhibitors. We expect that the present contribution will be helpful for the construction of stringent pharmacophore hypotheses of the selected targets which can be exploited as an efficient pharmaceutical filter and a coherent inhibitor strategy.

**Acknowledgements** Anu Manhas and PCJ acknowledge Science and Engineering Research Board (SERB), Department of Science and Technology (DST) for project grant through grant number EMR/2016/003025. MY Lone acknowledges the University Grants Commission (UGC), Govt. of India for the financial assistance.

## Compliance with ethical standards

**Conflict of interest** The authors declared no competing interest.

## References

1. Snow RW, Guerra CA, Noor AM, Myint HY, Hay SI (2005) The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature* 434:214. <https://doi.org/10.1038/nature03342>
2. World Health Organisation- WHO (2016) World malaria report 2015. WHO. <http://www.who.int/malaria/publications/world-malaria-report-2015/report/en/>. Accessed 19 April 2016
3. World Health Organisation- WHO (2015) Investing to overcome the global impact of neglected tropical diseases: 3rd WHO report on neglected tropical diseases, vol 3. World Health Organization- WHO, Geneva
4. Bhatt S, Weiss D, Cameron E, Bisanzio D, Mappin B, Dalrymple U, Battle K, Moyes C, Henry A, Eckhoff P (2015) The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature* 526:207. <https://doi.org/10.1038/nature15535>
5. Cui L, Mharakurwa S, Ndiaye D, Rathod PK, Rosenthal PJ (2015) Antimalarial drug resistance: literature review and activities and findings of the ICEMR network. *Am J Trop Med Hyg* 93:57–68. <https://doi.org/10.4269/ajtmh.15-0007>
6. Paloque L, Ramadanani AP, Mercereau-Puijalon O, Augereau J-M, Benoit-Vical F (2016) *Plasmodium falciparum*: multifaceted resistance to artemisinins. *Malar J* 15:149. <https://doi.org/10.1186/s12936-016-1206-9>
7. Verlinden BK, Louw A, Birkholtz L-M (2016) Resisting resistance: is there a solution for malaria? *Expert Opin Drug Discov* 11:395–406. <https://doi.org/10.1517/17460441.2016.1154037>
8. Müller IB, Hyde JE (2010) Antimalarial drugs: modes of action and mechanisms of parasite resistance. *Future Microbiol* 5:1857–1873. <https://doi.org/10.2217/fmb.10.136>
9. World Health Organisation- WHO (2016) Malaria vaccine: WHO position paper- January 2016. *Wkly Epidemiol Rec* 91:33–52. [http://apps.who.int/iris/bitstream/handle/10665/254285/wer9104\\_33-52.pdf?sequence=1&isAllowed=y](http://apps.who.int/iris/bitstream/handle/10665/254285/wer9104_33-52.pdf?sequence=1&isAllowed=y). Accessed Jan 2016
10. Leelananda SP, Lindert S (2016) Computational methods in drug discovery. *Beilstein J Org Chem* 12:2694. <https://doi.org/10.3762/bjoc.12.267>
11. Langer T, Hoffmann RD (2006) Pharmacophores and Pharmacophore Searches, vol 32. WILEY-VCH Verlag GmbH & Co., KGaA, Weinheim
12. Ekins S, Mestres J, Testa B (2007) In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *Br J Pharmacol* 152:9–20. <https://doi.org/10.1038/sj.bjp.0707305>
13. Ekins S, Mestres J, Testa B (2007) In silico pharmacology for drug discovery: applications to targets and beyond. *Br J Pharmacol* 152:21–37. <https://doi.org/10.1038/sj.bjp.0707306>
14. Yang S-Y (2010) Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discov Today* 15:444–450. <https://doi.org/10.1016/j.drudis.2010.03.013>
15. Kurogi Y, Guner OF (2001) Pharmacophore modeling and three-dimensional database searching for drug design using catalyst. *Curr Med Chem* 8:1035–1055. <https://doi.org/10.2174/0929867013372481>
16. Guner O (2002) History and evolution of the pharmacophore concept in computer-aided drug design. *Curr Top Med Chem* 2:1321–1332. <https://doi.org/10.2174/1568026023392940>
17. Xiao Z, Varma S, Xiao Y-D, Tropsha A (2004) Modeling of p38 mitogen-activated protein kinase inhibitors using the Catalyst™ HypoGen and k-nearest neighbor QSAR methods. *J Mol Graph Model* 23:129–138. <https://doi.org/10.1016/j.jmgm.2004.05.001>
18. Kirchmair J, Laggner C, Wolber G, Langer T (2005) Comparative analysis of protein-bound ligand conformations with respect to catalyst's conformational space subsampling algorithms. *J Chem Inf Model* 45:422–430. <https://doi.org/10.1021/ci0497531>
19. Kirchmair J, Wolber G, Laggner C, Langer T (2006) Comparative performance assessment of the conformational model generators omega and catalyst: a large-scale survey on the retrieval of protein-bound ligand conformations. *J Chem Inf Model* 46:1848–1861. <https://doi.org/10.1021/ci060084g>
20. Kirchmair J, Ristic S, Eder K, Markt P, Wolber G, Laggner C, Langer T (2007) Fast and efficient in silico 3D screening: toward maximum computational efficiency of pharmacophore-based and shape-based approaches. *J Chem Inf Model* 47:2182–2196. <https://doi.org/10.1021/ci700024q>
21. Kristam R, Gillet VJ, Lewis RA, Thorner D (2005) Comparison of conformational analysis techniques to generate pharmacophore hypotheses using catalyst. *J Chem Inf Model* 45:461–476. <https://doi.org/10.1021/ci049731z>
22. Manhas A, Patel A, Lone MY, Jha PK, Jha PC (2018) Identification of PfENR inhibitors: a hybrid structure based approach in conjunction with molecular dynamics simulations. *J Cell Biochem*. <https://doi.org/10.1002/jcb.27075>
23. Manhas A, Kumar SP, Jha PC (2016) Molecular modeling of *Plasmodium falciparum* peptide deformylase and structure-based pharmacophore screening for inhibitors. *RSC Adv* 6:29466–29485. <https://doi.org/10.1039/C6RA01071G>
24. Manhas A, Lone MY, Jha PC (2017) Multicomplex-based pharmacophore modeling coupled with molecular dynamics simulations: an efficient strategy for the identification of novel inhibitors of PfD-

- HODH. *J Mol Graph Model* 75:413–423. <https://doi.org/10.1016/j.jmglm.2017.04.025>
25. Lone MY, Kumar SP, Athar M, Jha PC (2018) Exploration of *Mycobacterium tuberculosis* structural proteome: an in silico approach. *J Theor Biol* 439:14–23. <https://doi.org/10.1016/j.jtbi.2017.11.021>
  26. Lone MY, Athar M, Gupta VK, Jha PC (2017) Prioritization of natural compounds against mycobacterium tuberculosis 3-dehydroquinate dehydratase: a combined in silico and in vitro study. *Biochem Biophys Res Commun* 491:1105–1111. <https://doi.org/10.1016/j.bbrc.2017.08.020>
  27. Lone MY, Athar M, Gupta VK, Jha PC (2017) Identification of *Mycobacterium tuberculosis* enoyl-acyl carrier protein reductase inhibitors: a combined in silico and in vitro analysis. *J Mol Graph Model* 76:172–180. <https://doi.org/10.1016/j.jmglm.2017.07.005>
  28. Lone MY, Manhas A, Athar M, Jha PC (2017) Identification of InhA inhibitors: a combination of virtual screening, molecular dynamics simulations and quantum chemical studies. *J Biomol Struct Dyn*. <https://doi.org/10.1080/07391102.2017.1372313>
  29. Accelrys Discovery Studio version 4.0, Accelrys, San Diego, USA. <https://www.accelrys.com/products/collaborativescience/biovia-discovery-studio/>
  30. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK (2006) BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities. *Nucl Acids Res* 35:D198–D201. <https://doi.org/10.1093/nar/gkl1999>
  31. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B (2011) ChEMBL: a large-scale bioactivity database for drug discovery. *Nucl Acids Res* 40:1100–1107. <https://doi.org/10.1093/nar/gkr777>
  32. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 4:187–217. <https://doi.org/10.1002/jcc.540040211>
  33. Weinstein JN, Myers TG, O'Connor PM, Friend SH, Fornace AJ, Kohn KW, Fojo T, Bates SE, Rubinstein LV, Anderson NL (1997) An information-intensive approach to the molecular pharmacology of cancer. *Science* 275:343–349. <https://doi.org/10.1126/science.275.5298.343>
  34. Zhou L, Griffith R, Gaeta B (2014) Combining spatial and chemical information for clustering pharmacophores. *BMC Bioinformatics* 15:S5. <https://doi.org/10.1186/1471-2105-15-S16-S5>