Comparing Drug Images and Repurposing Drugs with BioGPS and FLAPdock: The Thymidylate Synthase Case

Lydia Siragusa, Rosaria Luciani, Chiara Borsari, Stefania Ferrari, Maria Paola Costi, Gabriele Cruciani, Francesca Spyrakis. Comparing drugs image and repurposing drugs with BioGPS and FLAPdock: the Thymidylate Synthase Case. CHEMMEDCHEM. 11(15) pp: 1653-1666.
DOI: 10.1002/cmdc.201600121

Comparing drugs image and repurposing drugs with BioGPS and FLAPdock: the Thymidylate Synthase Case

Lydia Siragusa [1], Rosaria Luciani [2], Chiara Borsari [2], Stefania Ferrari [2], Maria Paola Costi [2], Gabriele Cruciani [3], Francesca Spyrakis [2,4,]*

[1] Molecular Discovery Limited, 215 Marsh Road, Pinner Middlesex-London HA5-5NE, United Kingdom; [2] Department of Life Sciences, University of Modena and Reggio Emilia, Via Campi 103, 41125, Modena, Italy; [3] Department of Chemistry, Biology and Biotechnology, University of Perugia, Via Elce di Sotto 8, 06123, Perugia, Italy; [4] current address: Department of Food Science, University of Parma, Viale delle Scienze 95/A, 43124, Parma, Italy.

*** Corresponding author**

Francesca Spyrakis

Department of Life Sciences, University of Modena and Reggio Emilia, Via Campi 103, 41125, Modena, Italy.

current address: Department of Food Science, University of Parma, Viale delle Scienze 95/A, 43124, Parma, Italy.

phone: 0039 0521 905669

e-mail: francesca.spyrakis@unipr.it

ABSTRACT

Repurposing and repositioning drugs has become an often pursued and successful strategy in the current era, where new medicals are more and more difficult to be found and approved.

We report an integrated BioGPS/FLAPdock pipeline for rapid and effective off-targets identification and drug repurposing. Our method is simply based on the structural and chemical properties of protein binding sites, that is, the ligand image, encoded in the GRID Molecular Interaction Fields. Proteins similarity is disclosed by the BioGPS algorithm by measuring the pockets MIFs overlapping, according to which pockets are clustered. Co-crystallized and known ligands can be cross-docked among similar targets, selected for binding *in vitro* experiments, and possibly improved for inhibition potency.

We used human Thymidylate Synthase as test case and searched the whole PDB for similar target pockets. We chose casein kinase II α as control and tested a series of its inhibitors against the TS template. Ellagic acid and apigenin were identified as TS inhibitors and different flavonoids were selected and synthesized in a second round selection. The compounds demonstrated to be active in the low micromolar range.

INTRODUCTION

In the threat and alarm of future drugs shortage, techniques aimed at drug repositioning and polypharmacology rationalization are becoming more and more popular. Antimicrobials are facing worrisome failures due to the spread and proliferation of multidrug resistant bacteria.[1-3] Cancer research has to deal with resistance to chemotherapy and molecular targeted therapies [4] and, in general, drugs are withdrawn from the market or during clinical phases because of toxic effects. In this scenario, so faraway from Ehrlich's magic bullet,[5] the development and discovery of drugs able to bind more than one single target, acting as a magic shotgun,[6] represent the main trend. Polypharmacology has emerged as the next paradigm of drug discovery,[7-12] and off-targets prediction is a key issue in bioinformatics and drug design. The identification of unknown targets not only can lead to unexpected applications or to multi-target drugs. It could also prevent adverse effects from occurring in clinical trials, or worse, when drugs have already reached the market. As well, drug repurposing, that is, using already approved and safe drugs for new targets, would spare a huge amount of money, time and resources. A number of methodologies have appeared in recent years with the aim of identifying off-targets, predicting side-effects and possibly find new applications for already known molecules.

Notable efforts have been made to develop *in vitro* assays and standardize procedure to determine the pharmacological profile of drug candidates. Even if experimental techniques can provide robust information, assays remain challenging and costly,[13] and computational methods currently represent a valuable strategy to be pursued in combination to *in vitro* analyses. Different methodologies have been developed for off-targets identification and drug repositioning.[14,15] While first approaches were based on sequence comparison,[16-20] in the last decade more exhaustive compound-based approaches were developed, starting from the assumption that similar chemicals should be able to bind similar pockets. Statistical and canonical correlations analyses were thus applied to link the ligand chemical space to the targets and to the possible related side-

effects (SEs).[11,21-25] At the same time, a number of phenotypic- and pathway-based methods were released, combining drug-disease relationships, clinically known SEs, gene-disease-drug connections and drug-drug interactions, in a knowledge-based perspective.[26-30] Also, different databases were constructed, and are currently available, to detect and predict relationships between drugs, target, side-effects and biological pathways.[31-33] More recently, approaches combining both chemical and target information have been reported showing that drug action is often unspecific, and underling the necessity of combining biology and chemistry to provide reliable molecular explanations for complex SEs.[34]

Approaches directly comparing protein pockets were also recently proposed, based on the assumption that similar binding sites can be targeted by similar ligands,[35-39] and that the structural and chemical information encoded into the binding pockets guide the recognition between macromolecules and ligands.[40-43] SiteEngine[41] Cavbase,[40] and FuzCav,[42] for instance, represent each pocket residue as a series of pseudocenters, encoding the physico-chemical properties essential for molecular interactions. These rule-based methods generally produce models, in which all pseudocenters are equally considered, without paying attention to the residues environment.

In this mazy plethora of options, structural-based methods could provide mechanistic indications on the off-targets selection and on the related occurrence of side-effects. Moreover, the possibility of finely tuning the effect of an approved or of a candidate drug towards a new target often relies on the modulation of their interactions. Structural information is, in fact, essential to understand and ameliorate the interaction of a compound with the binding site of a potential new target.

Trying to simplify as much as possible the off-target search, and going back to the chemical and physical principles of protein-ligand interaction, we developed the BioGPS algorithm, [44] based on the chemical/structural comparison of protein binding sites.[44,45] BioGPS (Global Positioning System in Biological Space) represents and compares pockets according to their “ligand image”

and not by any other rule-based residue feature.[40] Pockets are described by their Molecular Interaction Fields (MIFs, calculated by GRID [46,47]), that is, the shape, the hydrophobic regions, the H-bond donor and acceptor hot-spots a ligand would encounter upon entering the cavity. The BioGPS similarity score quantifies the geometrical and chemical similarity of multiple pockets upon alignment of their corresponding MIFs, and gives valuable clues about the structural correlation of proteins, even when belonging to distant and diverse families. Only GRID MIFs and multivariate statistical analysis are used to compare and cluster protein families.[48] No sequence-related information, ligand similarity or side-effects relationship is needed. With respect to other methods based on pseudocenters, which only represent the hydrophobic or H-bond donor/acceptor nature of the residues lining a cavity, BioGPS considers the MIFs generated by those residues according to their environment and, thus, the energetics of a pocket. The extension of a MIF depends on the generating groups and on the energy associated to a possible interaction with those groups. Larger is the MIF produced by one or more residue, higher is the probability to find a complementary group in that region, that is, an hydrophobic group if the pocket residue was hydrophobic or a H-bond donor or acceptor moiety if the residue was bearing a H-bond acceptor or donor side-chain. Comparing the MIFs means comparing the chemical and geometrical properties as well as the encoded energetics. Recent applications demonstrated BioGPS capability of predict off-target effects, classify protein families, justify polypharmacology, and rationalise selectivity between sub-families.[44,45,48]

Here we propose a specific pipeline (Figure 1) for investigating the biological space around a given target, identifying off-targets and, eventually, repurposing known drugs. The integrated approach includes a first BioGPS virtual screening (VS) step for the selection of the most similar pockets to the template within the Protein Data Bank, and then the docking of the ligands co-crystallized with the off-targets within the template binding site, using the FLAPdock algorithm implemented in FLAP.[49]

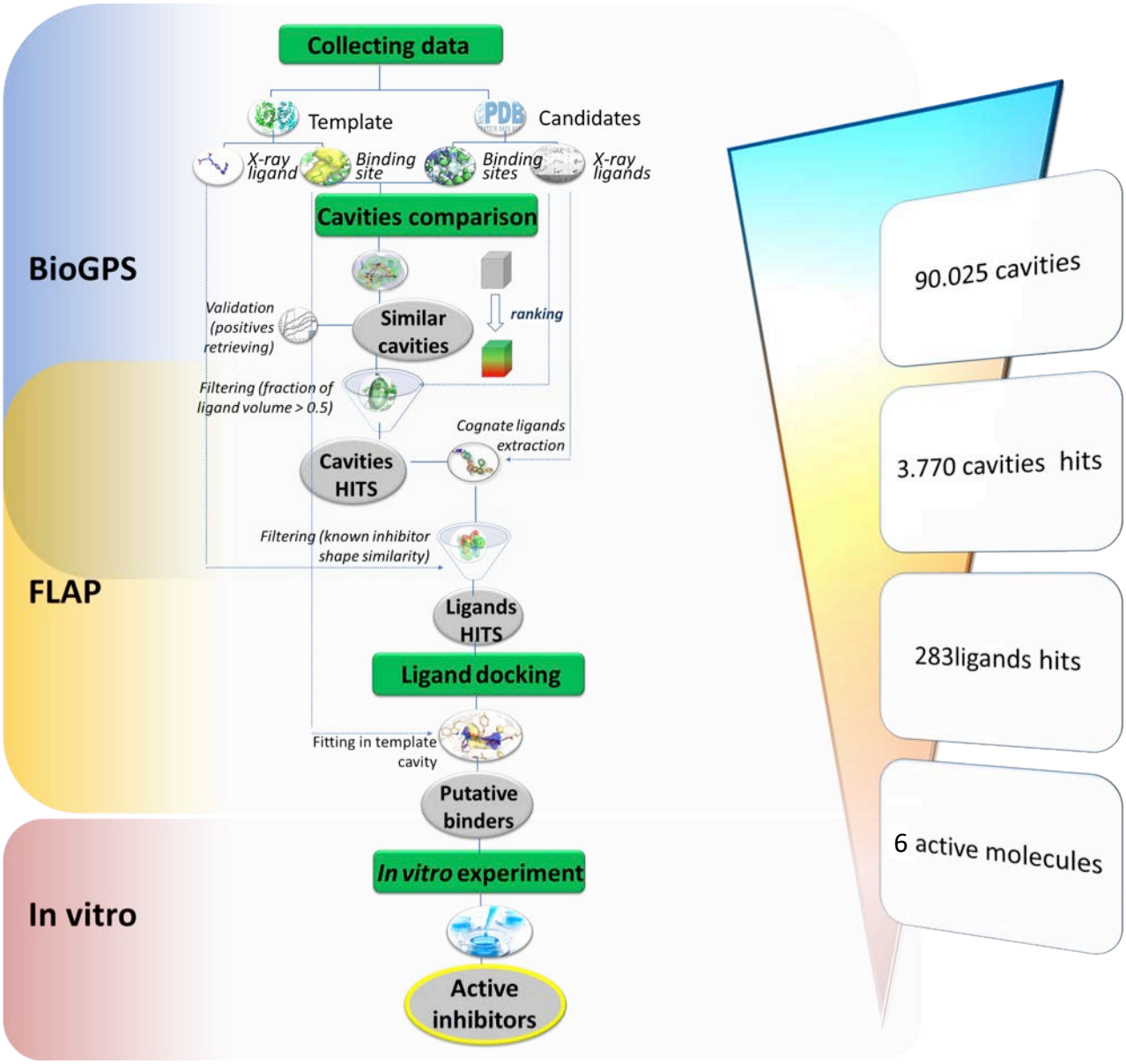


Figure 1. Workflow of the drug-repurposing approach based on the BioGPS/FLAPdock integrated technology.

The pipeline is composed of four main steps, i) data collection, ii) cavities comparison and selection, iii) ligand docking, iv) *in vitro* binding experiments.

i) *Data collection.* The first step consists in the selection of the protein template three-dimensional structures and of the database of proteins to compare to it. For each protein structure the co-crystallized ligand and the cavity containing it, the binding site, are detected.

ii) *Cavities comparison.* The template pocket is compared to all the other pockets by using the BioGPS algorithm. Cavities in the database are ranked accordingly to their similarity with the template. The algorithm capability of retrieving pockets belonging to the same template protein family is evaluated by means of enrichment analyses. An established similarity threshold (Global Product > 0.7) is used for selecting similar cavities. In order to avoid sites not completely occupied by ligands, only cavities containing at least 50% of the co-crystallized ligand volume are retained. Once cavities hits are selected, co-crystallized ligands are extracted and filtered accordingly to their volume similarity with respect to known template inhibitors.

iii) *Ligand docking*. The extracted ligands are docked within the template binding site with the FLAPdock algorithm. Most promising ligands are selected according to the FLAP S-Score value, to their pseudo MIFs complementarity with the pocket MIFs, and to the number of hydrogen bonds formed with the residues lining the cavity.

iv) *In vitro binding experiments*. Selected ligands are tested *in vitro* for TS inhibition.

We used as template the human Thymidylate Synthase (TS). Human and bacterial TS catalyses the reductive methylation of 2'-deoxyuridine-5'-monophosphate (dUMP) to 2'-deoxythymidine-5'-monophosphate (dTMP), using 5,10-methylenetetrahydrofolate as one-carbon methyl donor (mTHF). The reaction evolves through the formation of a covalent bond between dUMP and the catalytic Cys195, the entrance of mTHF into the binding site and the transfer of a methylic group on dUMP, thus transformed in dTMP.[50] After its release, dTMP is phosphorylated by two successive steps to 2'-deoxythymidine-5'-triphosphate (dTTP), an essential precursor for DNA synthesis. This pathway is the sole intracellular de novo source of dTTP. It derives that human TS represents a good pharmacological target for anticancers and antimicrobics. Nucleotide- and folate-like inhibitors are, in fact, used in cancer chemotherapy because of the cytotoxic effects of

thymidylate depletion.[51-53] As well, microbial TSs demonstrated to be suitable targets for antimicrobic agents.[54,55]

TS binding site was used as template to screen the whole PDB database, looking for similar cavities. Human and bacterial TS were reasonably identified as the most similar proteins, being followed by apparently diverse candidates as kinases, proteases, phosphodiesterases, nuclear receptors, and chaperons among the others. Statistical and network analyses were used to rationalize the investigated biological space and the connection, in terms of similarity, among the selected pockets. The results enclose the strength of this approach, able to automatically and quickly pick out similarity between the same and different protein families.

Looking for possible new TS ligands among known chemicals, ligands co-crystallized in the most similar pockets to the TS template were subsequently docked with the FLAPdock algorithm in TS cavity. Interestingly, the most promising molecules were inhibitors of casein kinase IIα (CKIIα), which emerged as a TS off-target. Two co-crystallized ligands and a series of related flavonoids were tested *in vitro* for inhibition activity towards TS. Five compounds demonstrated to inhibit TS in the low micromolar range, thus supporting the potential of the BioGPS approach for drug repurposing campaigns.

RESULTS AND DISCUSSION

Thymidylate Synthase can exist in different configurational states, i) apo inactive form; ii) inactive form complexed with peptides binding the homodimer interface; ii) active form complexed with dUMP; iii) active form complexed with dUMP and mTHF or an antifolate drug (ternary complex). TS undergos significant conformational rearrangements upon dUMP binding. In particular, in the native un-bound form the 181-197 loop, containing the catalytic cysteine, is rotated of about 180° with respect to the binary/ternary complex. Consequently, the catalytic Cys195 thiol group is 10 Å away from the active site, confirming the inactivity of the enzyme in this conformational state. On

the contrary, upon dUMP binding the enzyme assumes the closed active conformation.[56] Looking for new unknown TS ligands, we focused on the protein closed active conformation in presence of the dUMP substrate (Figure 2).

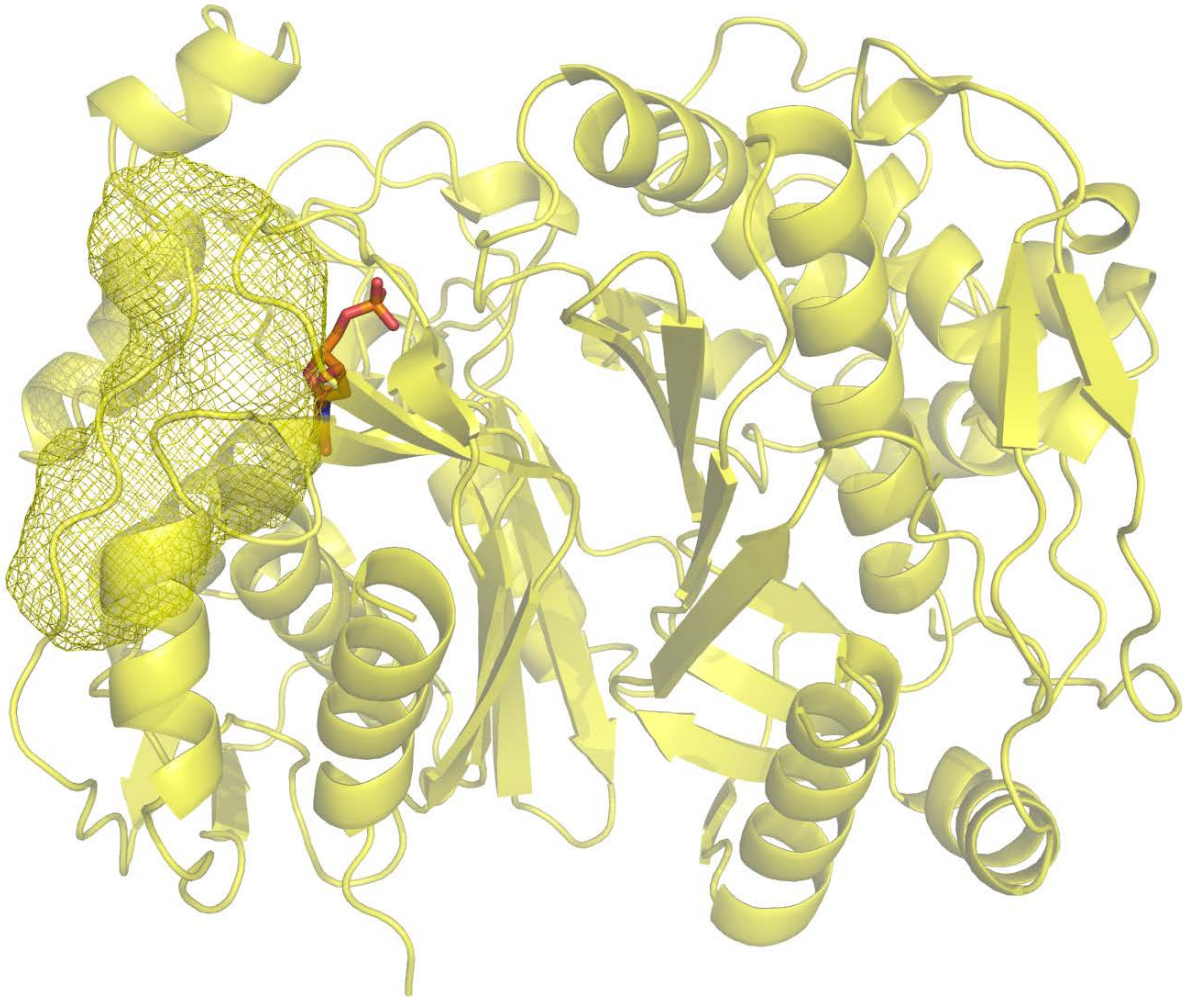


Figure 2. Cartoon representation of human TS. The binding site of one monomer is represented as yellow mesh lines, dUMP is shown as orange capped sticks.

In the PDB many different apo forms of wild type human TS have been deposited, there are a few binary complexes of the inactive form co-crystallized with peptides binding at the homodimer

interface, while there are no binary complexes for the active form with the dUMP substrate. Only four ternary complexes, i.e. 1hvy (hTS+dUMP+raltitrexed), 1i00 (hTS+dUMP+raltitrexed); 1ju6 (hTS+dUMP+LY231514); 1juj (hTS+dUMP+LY231514) are present. According to the resolution and the backbone completeness, 1hvy was selected for modelling TS binding site.[56] The antifolate raltitrexed was removed from the crystallographic structure, leaving only the dUMP and thus limiting the search to the pocket normally occupied by mTHF or other antifolate drugs. A careful structural and energetic analysis of binding site water molecules was also performed, to identify structural waters able to modify the cavity shape and the possible interaction with ligands. Waters 435 and 622, mediating the interaction of the protein with raltitrexed and dUMP, respectively, and thus contributing to the complex stabilization, were identified. The *Fixpdb* tool implemented in BioGPS was used to calculate the free energy of the aforementioned waters within the binding pocket. The two selected molecules showed a free energy value lower than -8 kcal/mol, a strong evidence of their importance in complex formation and stabilization. Thus, two different human TS structures were finally modelled, i) TS with dUMP and without any water molecule (TSp); ii) TS with dUMP and two bridging water molecules (wat435 and wat622; TSpw). The corresponding pockets were automatically identified by using the FLAPsite tool and used as template to perform the following screening and docking experiments, in the perspective that the incoming ligands might displace any water (TSp) or retain and exploit the existing ones (TSpw).

1st step: screening the pockets.

The two TS pockets were used as templates to screen the PDB with BioGPS. In particular, 90.025 pockets were screened, which correspond to all the protein cavities present in the PDB and co-crystallized with a ligand at the moment the database was downloaded (September 2014).

BioGPS superposes cavities by aligning their MIFs. Template MIFs were compared to cavity database MIFs, calculating for each pair-wise comparison a set of nineteen FLAP Scores, representing the similarity of the match (see the Materials and Methods section for further details on BioGPS and FLAP). In particular, the Global Product (GlobP), produced by multiplying all the FLAP Scores, was used to evaluate the similarity degree between the templates and the candidate cavities. The GlobP score ranges from 0, for null superposition, to 1, for a complete pocket overlapping, and provides a global evaluation of both geometrical and chemical similarity. Thus, the 90.025 candidate cavities were ranked according to the GlobP score for both TSp and TSpw. The distributions of the GlobP scores for TSp and TSpw are reported in Figure 3a, in purple and cyan respectively.

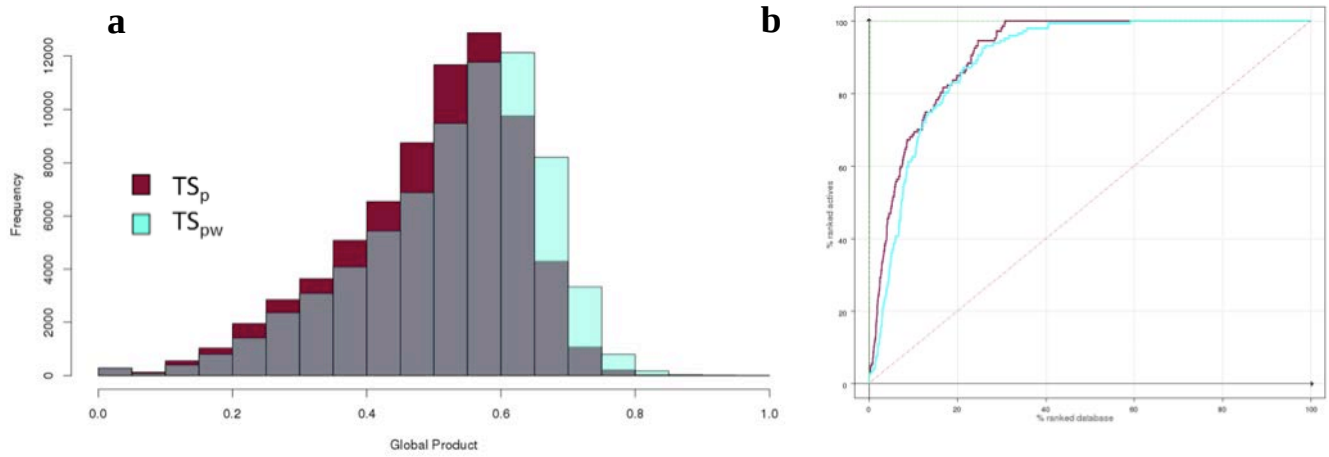


Figure 3. BioGPS virtual screening statistics. a. Distribution of the similarity Global Product score calculated by the BioGPS virtual screening**.** TSp and TSpw, coloured red and cyan respectively, were used as template pockets. Overlap areas are grey coloured. **b.** Enrichment curves calculated for the two virtual screenings. All the pockets in the dataset belonging to TS were considered actives, all the other pockets inactives. The curves identifying the VSs performed using TSp and TSpw as templates, respectively, are colored purple and cyan.

The plot suggests that the water molecules enhance the promiscuity of the cavity. Indeed, using TSpw as template, the algorithm identifies a higher number of cavities with higher similarity, that is, having higher GlobP (grey and cyan bars). On the contrary, when TSp is used as template fewer cavities presented such high GlobP value. To demonstrate the robustness of the procedure, the GlobP score was used to evaluate the BioGPS performance in retrieving protein binding sites that belong to the same family of the query (259 TS pockets over 90.025 pockets). The PDB entries of thymidylate synthases were used to enrich the remaining cavities in the database and the GlobP was used again as the ranking score. Figure 3b reports the enrichment curves obtained for each of the two templates. In both cases about 80% of TS binding sites were retrieved within the first 20% of the screened database, confirming the strength of the BioGPS approach.

To select the most similar cavities to the TS, a GlobP threshold equal to 0.7 was set, according to previous analysis[44]. 1.336 cavities, related to 606 unique proteins, and 4.349 cavities, related to 1.513 unique proteins, were selected as hits for TSp and TSpw, respectively (Figure 4). 1.208 cavities (583 proteins) were selected as common hits for both templates. To consider both binding cases (water displacement [TSp] and water-mediated binding [Tspw]), we retained all the 4.476 identified pockets, corresponding to 1.536 proteins. To further reduce this large amount of cavities and focus on those most similar to the templates, we considered only binding sites containing a fraction of ligand volume (FV) > 0.5, that is, occupied by at least 50% of the ligand. This filtering was applied to discard pockets not completely occupied by the co-crystallized ligand. Supposing to have a cavity A similar to a cavity B, if cavity B contains only a fraction of its cognate ligand, the transferability of that ligand to cavity A might be misleading, because only a small part of the ligand is complementary to that pocket.[57,58] After applying the filtering procedure based on FV, we obtained overall 3.770 cavities and 1.297 proteins (Figure 4). These cavities contained 1.361 different ligands, being some of them present in different pockets. No relevant compounds as solvents, i.e. ethylene glycol, glycerol, or prosthetic groups were discarded. The remaining ligands

were compared, in terms of volume, to a set of known human TS inhibitors retrieved from the ChEMBL database (see Materials and Methods for further details). Those having a volume higher than the smallest TS ligand and lower than the biggest one were retained, to consider compounds in the same volume range of known TS inhibitors. We ended up with 283 ligands belonging to 135 proteins and 317 pockets (Table S1).

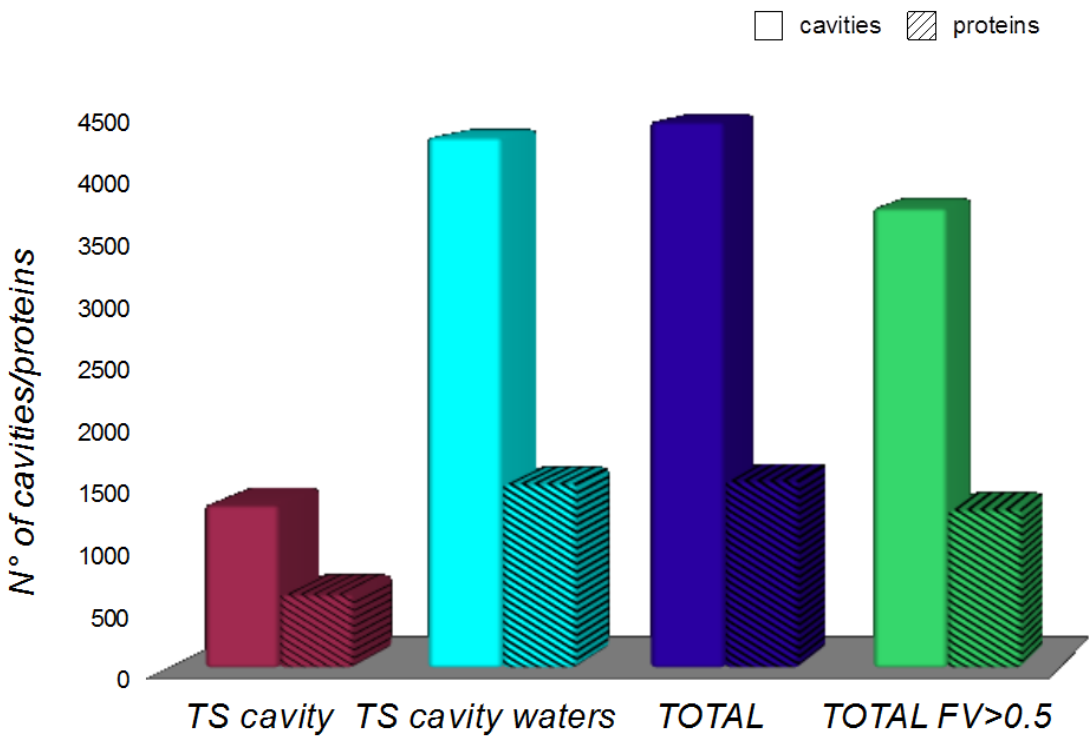


Figure 4. Cavity hits, and corresponding proteins, identified by the BioGPS VS. The hits identified for the TSp and the TSpw templates are respectively coloured purple and cyan. The total number of hits is coloured blue, while the green bars correspond to the hits number filtered for the ligand fraction volume.

Connecting the pocketome.

To determine how the selected pockets are connected and to identify how many unique pocket types we are dealing with, we performed a connectivity analyses on the 317 hits, by screening all the cavities towards themselves. We then built a binding site similarity network with Cytoscape,[59,60] where each node represents a pocket and pockets are connected and clustered if they have a GlobP higher than 0.4. Higher is the GlobP for a cavity pair, closer will be the two cavities in the graph, that is, more similar from a structural, chemical and energetic point of view. This lower threshold, with respect to the 0.7 value used for the initial VS, was used to find connections among all the selected hits and not only with the template. In Figure 5 higher thresholds corresponding to 0.8 and 0.95 were instead considered to connect TSp and TSpw to the others and to themselves, respectively. This was done to avoid over-connections between the original templates and the other pockets. Different protein classes are color-coded according to the legend. Pockets sharing no similarity with any other in the set are classified as singletons and are reported in the two lines under the network graph.

The high color/cluster correspondence reported in Figure 5 shows how BioGPS is not only able of detecting pocket similarities, but also of clustering proteins according to their binding site properties. In our study pockets are only described in terms of four basic molecular interaction fields, that is, the shape, the hydrophobicity, the H-bond donor and acceptor character. These simple properties, mixed up in the different FLAP scores and combined in the cumulative GlobP, are able to represent pockets to such an extent that proteins belonging to the same family are easily picked up and clustered together. No sequence or overall structure architecture is considered, which means that binding site similarity is enough to separate proteins in the corresponding families.

Chaperones and nuclear receptors (NRs) are the better clustered families, with no element falling out of the group (red and purple clusters, respectively). The homogeneity of these two groups,

including only HSP90 for chaperons and the vitamin D3 receptor and the estrogen receptor for the NRs class, certainly facilitates the tight clustering. Transferases and oxidoreductases are also well clustered (cyan and green spheres), with a few elements falling in other classes or classified as singletons. Kinases, coloured blue, fall all within the same cluster on the right of the graph, with only a few exceptions reported in the singleton section. The aspect of the kinase cluster is interestingly different with respect to the others, being somehow stretched and widen. We carefully inspected the cluster members, finding that most of them contain type I inhibitors and present the typical DFG-in conformation. Nevertheless, given the known kinase flexibility, some structure presents broken chains and different arrangement of the P- and activation loops, possibly responsible of slightly varied MIFs. Moreover, quite different kinases populate the cluster, as CDK2, CKIIα, Lyn kinase, Rho associated kinase, IRAK4, and others. Among the apparently spread hydrolases class (pink spheres) we can identify several consistent sub-clusters. The pocket group on the left of the network includes only adenosine deaminases, the following single-connected pink group phosphodiesterase (PDE) 4 and 10, while the 4-member cluster on the top of the graph includes three chitinases and one endochitinase. As well, also in the class defined as "others" (yellow spheres) we can identify sub-clusters. For instance the pockets clustered on the top and on the right-bottom of the graph all belong to integrin and to transcription factors, respectively, while those forming the first cluster in the singleton section correspond to isomerases. Notably, the two TS cavities lie close to each other and in a rather central position with respect to the main clusters. As mentioned, higher thresholds were considered to depict TS connectivity since TS would be connected to any pocket in the graph.

This analysis has the main advantage of depicting the connections among different protein classes, regardless the similarity they have with the original template, TS in the present case. In a drug repurposing perspective, for instance, we could try to exchange ligands among nuclear receptors and chaperons or nuclear receptors and oxidoreductases, while more difficultly among chaperons

and oxidoreductases. As well, chaperons ligands could be possibly repurposed for PDEs and *viceversa*, while it would be harder, in principle, to relocate PDEs ligands in nuclear receptors. In a polypharmacology perspective, pocketomes could be easily analyzed and investigated for multi-target therapies or for unpredicted and unknown side-effects. This underlines the versatility of the BioGPS approach and the extent of the possible related applications.

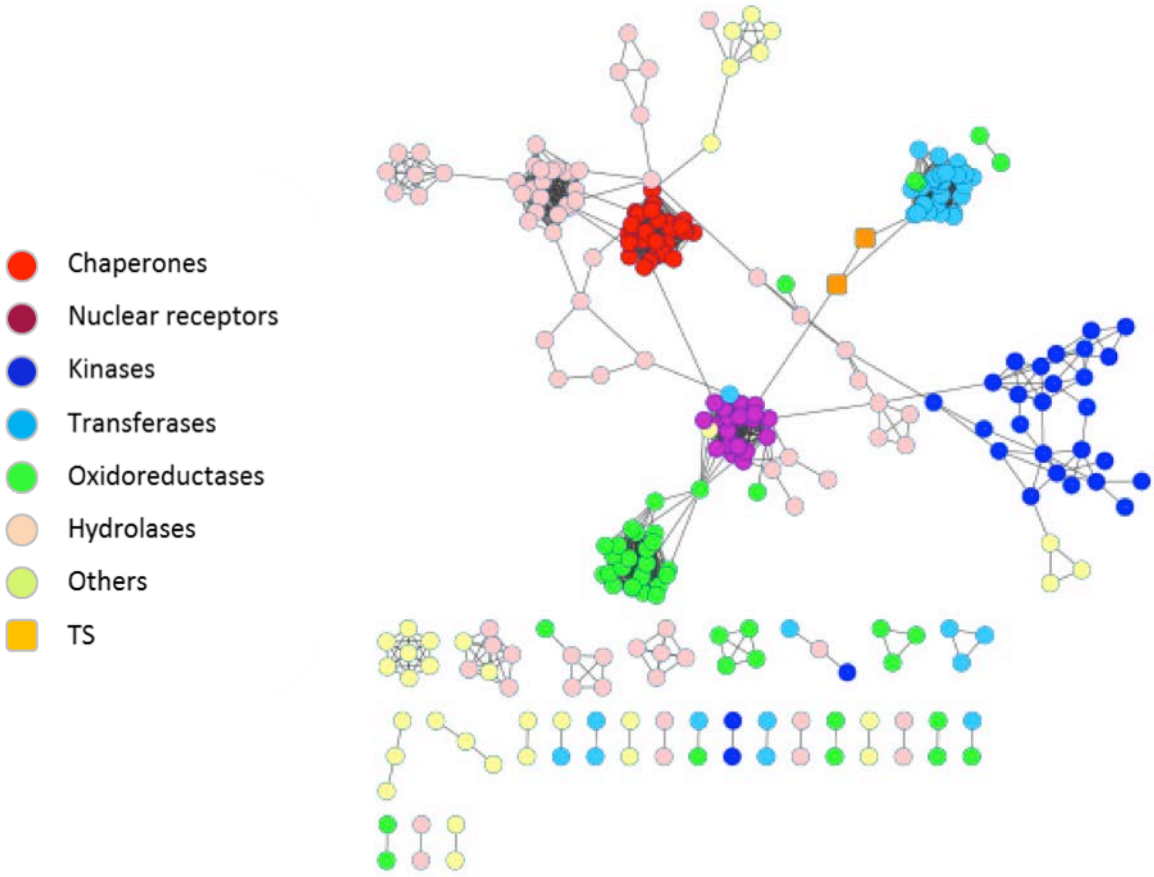


Figure 5. BioGPS-network. Similarity network of the 317 selected cavities built according to the cavities pair Globlal Product estimated by BioGPS. Each node represents a cavity and each edge between two nodes represents the similarity between them (shorter is the distance, higher is the

similarity). The different protein classes are colour-coded according to the legend. Singletons are reported below.

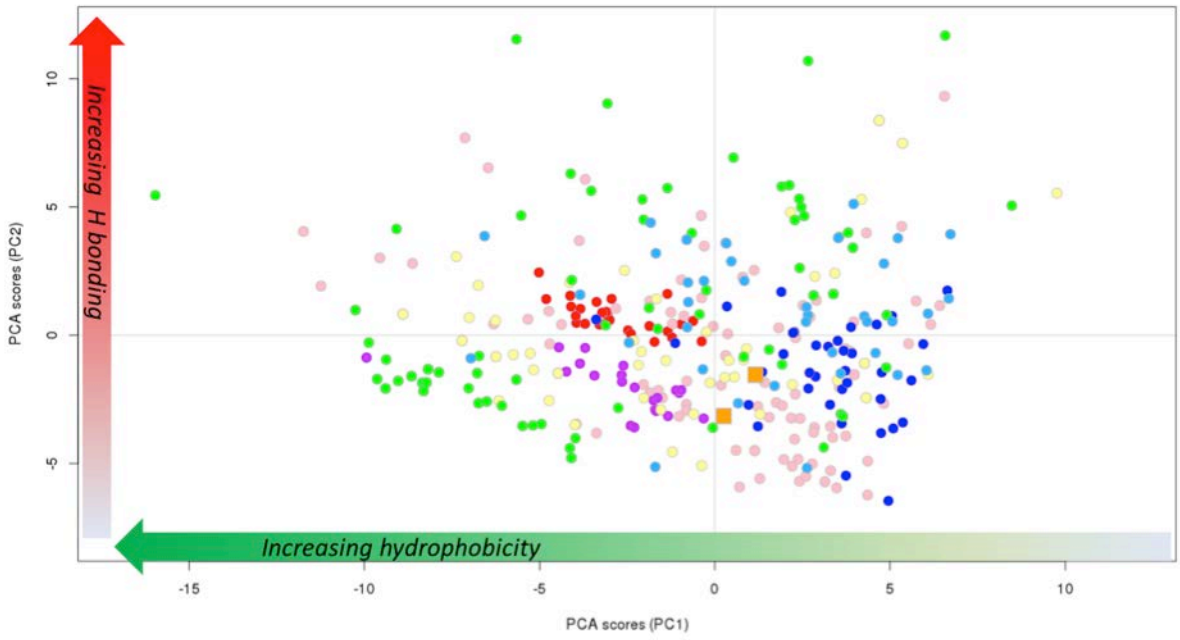
Rationalizing the similarity.

We then analysed the same 317 cavities with Volsurf, meant to characterize and separate entities according to their morphological and chemical properties.[61] Here the entities, the cavities, are described by the corresponding MIFs, then encoded into morphological and interaction energy descriptors. Descriptors are analyzed with statistics as principal component analysis, grouping the selected cavities on the basis of the same descriptors (Figure 6a). No comparison algorithm is used, that is, cavities are not aligned and compared. Besides, no three-dimensional information accounting for MIFs organization is provided. The first two components PC1 and PC2 explain the 37% and 19% of variance among the set, respectively, while PC3 accounts for only the 8%. We thus retained PC1 and PC2 sufficient to highlight the differences among the pockets. Again, chaperones and nuclear receptors are well clustered and close to each other, confirming the direct connection previously observed. Kinases and transferases groups are more spread and somehow overlapping, while the other classes basically explore the whole bidimensional space.

PC loadings measure the importance of each descriptor in accounting for the variability of the corresponding PC. The PC loadings plot (Figure S2) reveals that the hydrophobic descriptors are better able to differentiate cavities on the first components while H-bond (acceptor and donor) descriptors influence the cavities position on the second component. Since PC1 explains the higher percentage of variance, we can deduce that hydrophobic properties and hydrophobic interactions represent the main driving force that differentiates the classes. We thus expect chaperons and NRs to have a more lipophilic binding site with respect to kinases and transferases, while the remaining classes, being also more heterogeneously populated, present variable pockets, both in terms of hydrophobicity and H-bonding properties. PC2 clearly separates only chaperons

and NRs. The central position assumed by TS suggests that its pocket has average morphological and chemical properties with respect to the other cavities in the set. This is in agreement with the aim of the present work, that is, the identification of pockets similar to TS binding site. TS centrality is even more evident in the network graph reported in Figure 6b. Calculating the distance between each pair of objects in the bidimensional space, we obtained a square matrix, which reports the distance among different pockets in terms of hydrophobicity, polarity and morphology properties. The clusterization of chaperons and NRs is even more evident, with respect to the PCA score graphs (Figure 6a). Interestingly, kinases and transferases are almost completely superimposed, while in the BioGPS network (Figure 5) they are well clustered and separated. Again, closeness means similarity, but we have to consider that this similarity is only related to the MIFs volume and extension, and no information regarding their spatial organization is included. Proteins having a comparable hydrophobic/polar character, as kinases and transferases, could have a very different distribution of the molecular interaction fields within the cavity. It derives that the cognate drugs could not be so easily repositioned. On the contrary, closeness in the BioGPS network means MIFs similarity both in terms of volume and of three-dimensional organization. This supports again the BioGPS utility in a drug repurposing perspective, when we are looking for similar binding site and ligand “images”.

a


Increasing H bonding
Increasing hydrophobicity
PCA scores (PC2)
PCA scores (PC1)
10
5
0
-5
-15
-10
-5
0
5
10

b

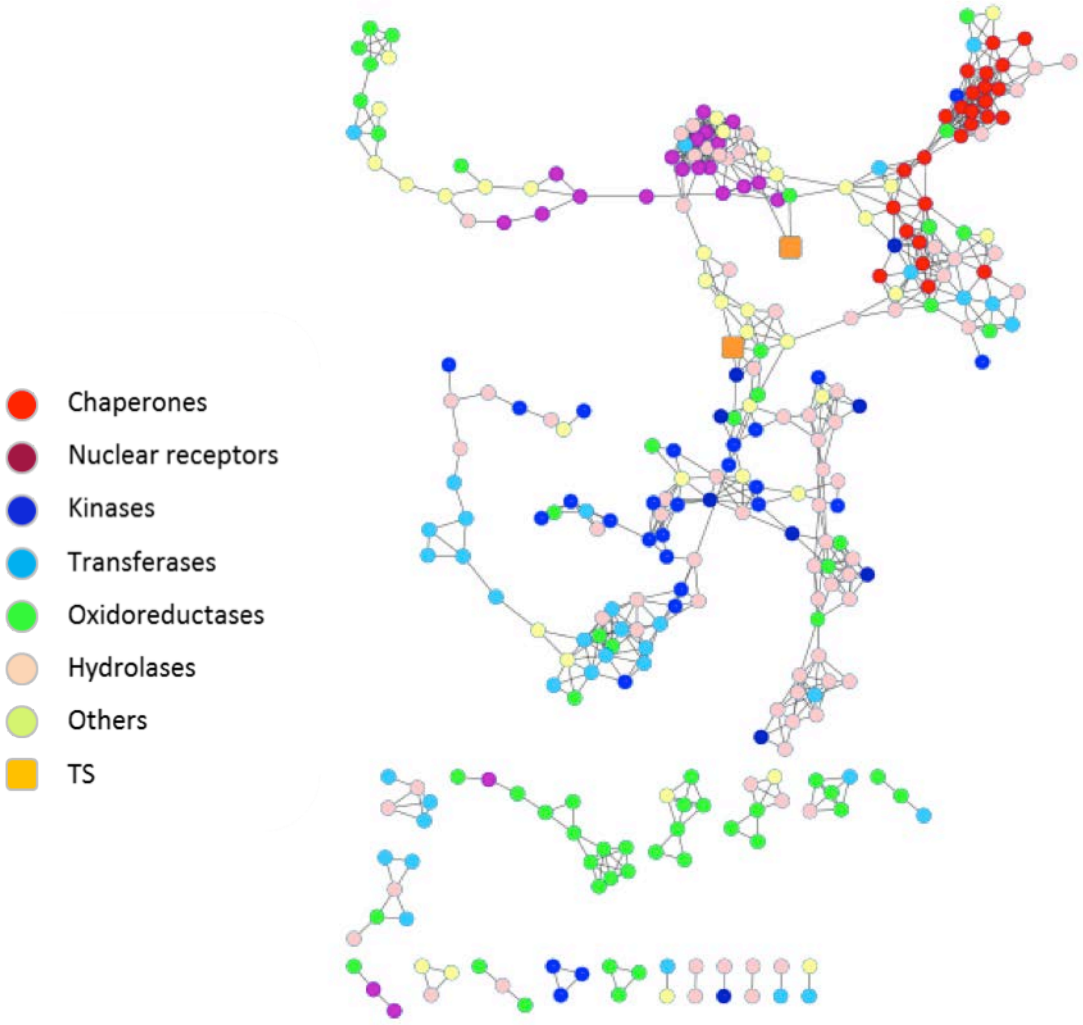

Chaperones
Nuclear receptors
Kinases
Transferases
Oxidoreductases
Hydrolases
Others
TS

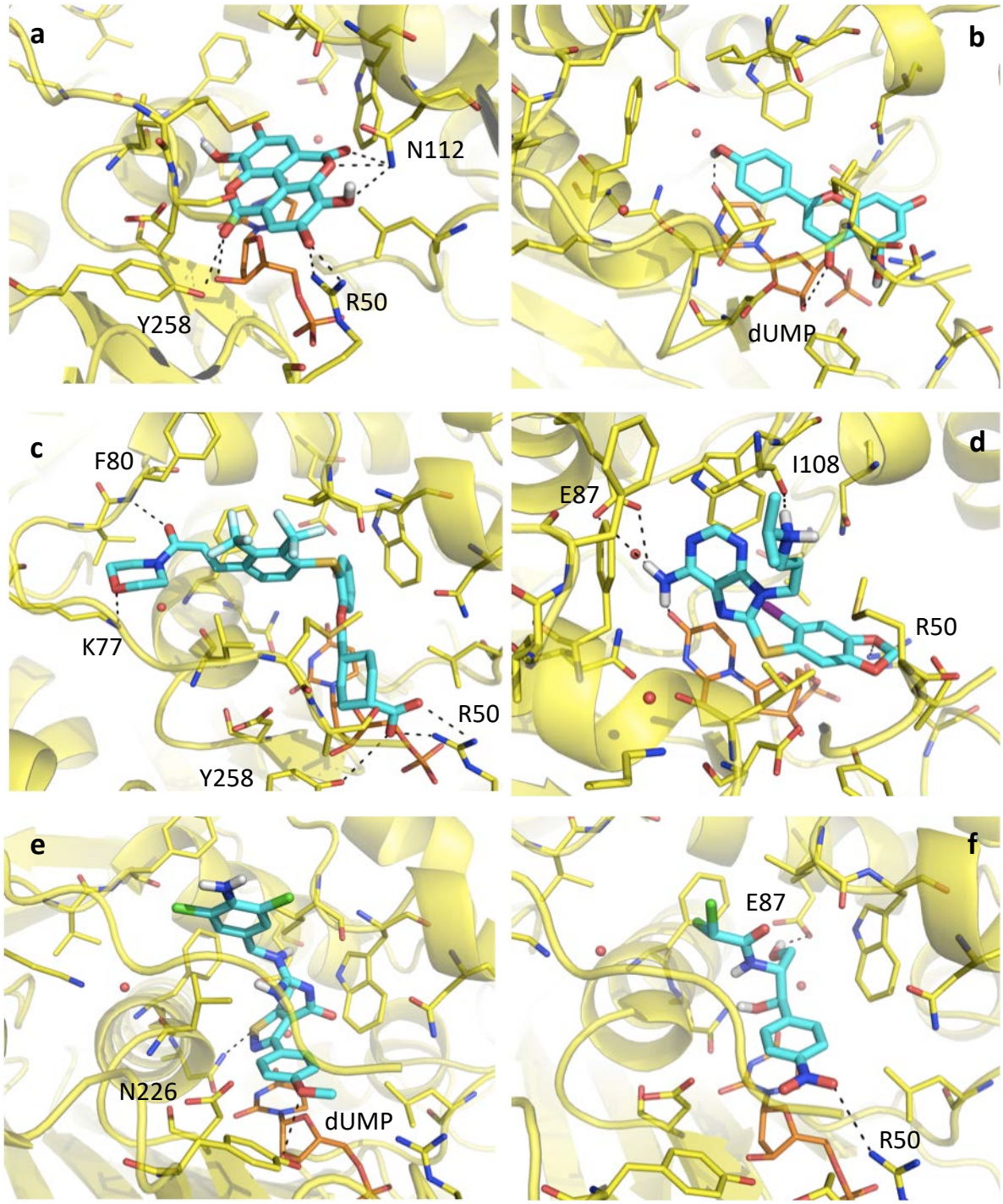
Figure 6. a. PCA scores plot of the Volsurf descriptors calculated for the 317 selected cavities. Cavities are positioned in the bi-dimensional space in order to maximize the variance among the component scores. The first component (PC1) explains the variance in terms of hydrophobicity (GRID probe = DRY), the second component (PC2) in terms of hydrogen bonding capability (GRID probes = N1, O). **b. Volsurf-network.** Similarity network of the 317 selected cavities built according to the Volsurf descriptors calculated for each of them. Each node represents a cavity and each edge between two cavities the distance in terms of physico-chemical properties. Distances are calculated from the scores assigned in the PCA. Cavities sharing physical-chemical properties (hydrophobicity, polarity, morphology) are connected by short edges. The different protein classes are colour-coded according to the legend. Singletons are reported below.

2nd step: docking cognate ligands.

From the 317 cavities we retrieved 283 co-crystallized different ligands (Table S1). The ligand dataset was implemented with tautomers and protomers and docked within TSp and TSpw with FLAPdock, the docking tool implemented into FLAP.[49] Compounds were first filtered using the FLAP S-Score and retaining all molecules with a score value higher than 0.90, according to previous observations [data not shown]. The selected compounds were visually inspected and further filtered according to the complementarity of the pocket MIFs with the ligand pseudoMIFs, and to the number of hydrogen bonds formed with the residues lining the cavity.[62] The ten most promising molecules were selected and are reported in Figure 7 and Table 1. Ligands were retrieved from casein kinase II α, LFA-1 integrin, heat shock protein 90, beta-secretase 1, chloramphenicol acetyltransferase 3, serine/threonine-protein kinase/ endoribonuclease IRE1 and Rho associated protein kinase.

The ligands, quite different one from each other in terms of shape, volume and chemical properties, present, predictably, diverse interactions with the pocket residues. Only the pi-pi contact with dUMP is maintained overall, with the exception of E2M ligand (Figure 7c). Among the ten selected compounds, ellagic acid and apigenin presented the most promising docking pose. In particular, ellagic acid makes hydrogen bonds with Arg50, Asn112 and Tyr258, while apigenin is H-bonded to a pyrimidine carbonyl and to the hydroxyl group of the deoxy-ribose dUMP. As well, slight different apigenin poses within TS pockets allow the formation of contacts with the aforementioned residues, thus underlining the plasticity of the complex and the possibility of making several and different interactions. Both ligands are involved in a pi-pi interaction with the dUMP pyrimidine moiety (Figures 7a,b). The other molecules also mainly contact Arg50, Asn112, Tyr258 and dUMP. Additional H-bonds are formed with Phe80, Glu87, Ile108, Asp218, Asn226, Ala312 (Figure 7).

Interestingly, both ellagic acid and apigenin are known inhibitors of casein kinase IIα (CKIIα). CKII is an attractive anti-neoplastic and antiviral target, essential for cell viability and with a plethora of cellular targets.[63] The catalytic subunits of CKII, α and α', are constitutively active, either alone or in combination with the beta subunit, a necessary property for the continuous need to phosphorylate its numerous targets, but also potentially dangerous in neoplastic pathologies and viral infections.[64]


a
N112
R50
Y258
b
dUMP
c
F80
K77
R50
Y258
d
E87
I108
R50
e
N226
dUMP
f
E87
R50

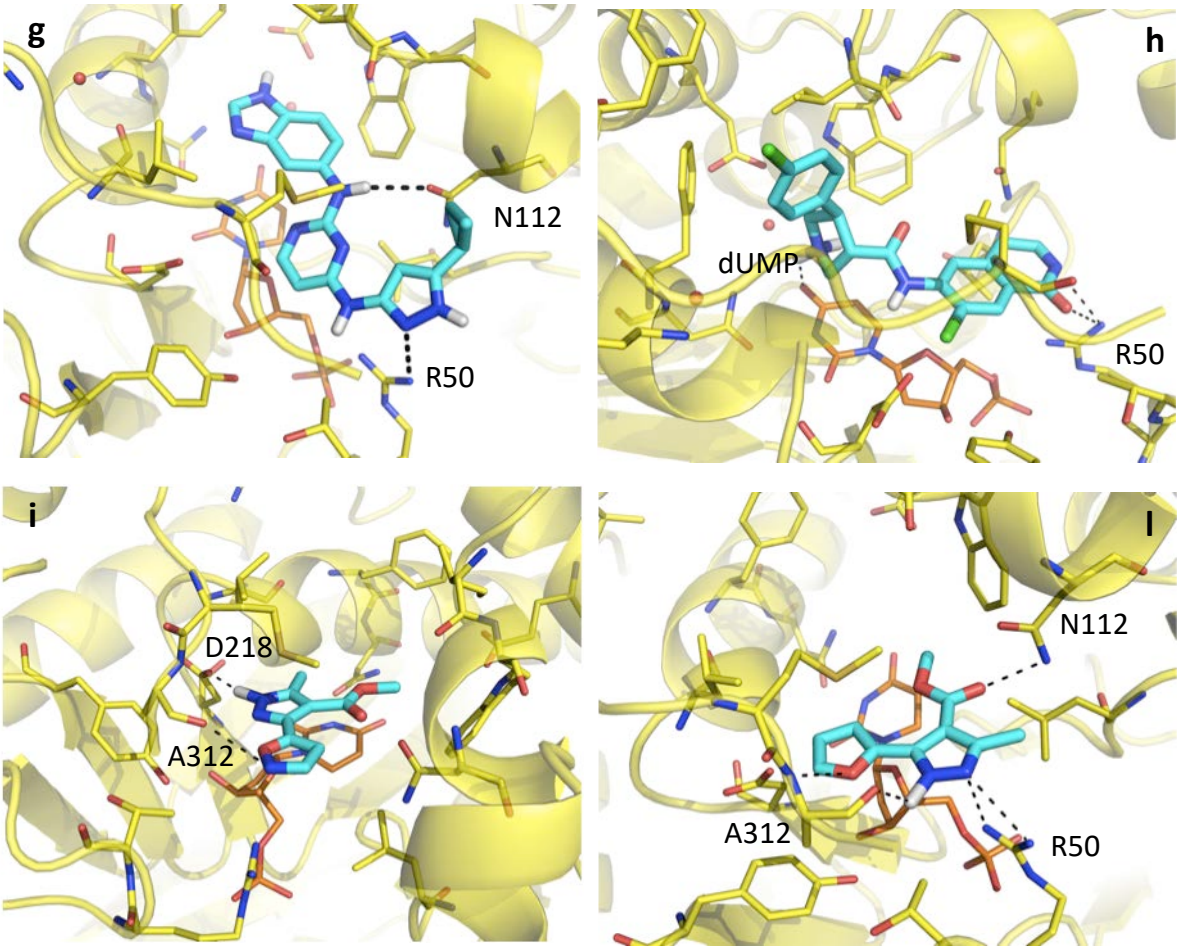


Figure 7. Docking poses of the best ten selected compounds within hTS binding pockets as predicted by FLAPdock. The PDB ligand code and the ligand name are reported as follows **a.** REF; ellagic acid; 2,3,7,8-tetrahydroxychromeno[5,4,3-cde]chromene-5,10-dione. **b.** AGI; apigenin; 5,7-dihydroxy-2-(4-hydroxyphenyl)-4H-chromen-4-one. **c.** E2M; cis-4-{[2-({4-[(1E)-3-morpholin-4-yl-3-oxoprop-1-en-1-yl]-2,3-bis(trifluoromethyl)phenyl}sulfanyl)phenoxy]methyl}cyclohexanecarboxylic acid **d.** H71; 8-[(6-iodo-1,3-benzodioxol-5-yl)thio]-9-[3-(isopropylamino)propyl]-9H-purin-6-amine. **e.** 0VA; N-[N-(4-amino-3,5-dichlorobenzyl)carbamimidoyl]-3-(4-methoxyphenyl)-5-methyl-1,2-thiazole-4-carboxamide. **f.** CLM; chloramphenicol; 2,2-dichloro-N-[(1R,2R)-2-hydroxy-1-(hydroxymethyl-2-(4-nitrophenyl)ethyl]aceta-mide. **g.** APJ; N~2~-1H-benzimidazol-5-yl-N~4~-(3-cyclopropyl-

1H-pyrazol-5-yl)pyrimidine-2,4-diamine. **h.** 3ND; (3S,4R)-N-(7-chloro-1-oxo-1,4-dihydroisoquinolin-6-yl)-4-(4-chlorophenyl)pyrrolidine-3-carboxamide . **i.** 2D3; ethyl 3-isoxazol-5-yl-5-methyl-1H-pyrazole-4-carboxylate. **l.** 37D; methyl 5-furan-2-yl-3-methyl-1H-pyrazole-4-carboxylate. The dUMP is shown in orange color code sticks. Residues involved in H-bonding the ligands are labeled. When present, water molecules are displayed as red spheres. Complexes **a-h** were obtained by docking in TSp, while complexes **i,l** by docking in TSpw.

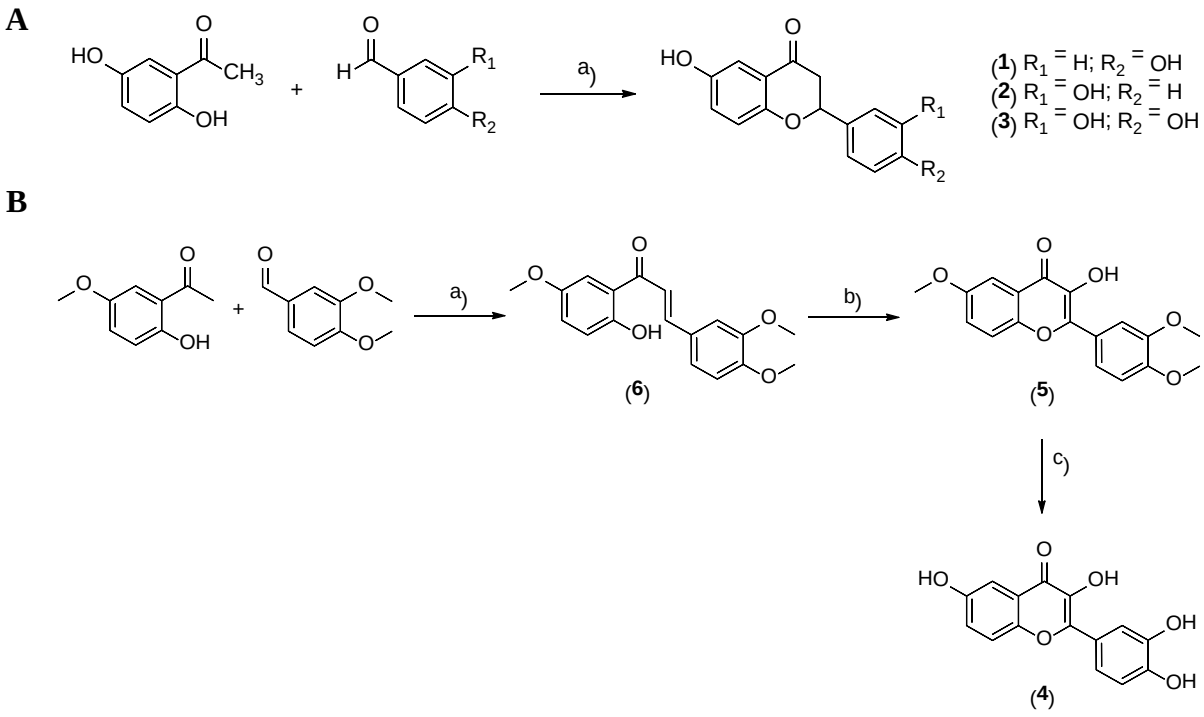
Biological evaluation of ligands and synthesis of second round compounds.

As mentioned, ellagic acid and apigenin are the most interesting compounds, and are both inhibitors of CKIIα. We thus experimentally evaluated them for inhibition towards our template TS and performed enzyme kinetics on both of them.

TS is a double substrate enzyme and competition with respect to dUMP or with respect to the folate cofactor can be performed. Following the computational model, dUMP was considered the fixed substrate and used at saturating concentration, while the folate cofactor was used as the limiting substrate for the competition kinetic. Results are reported in Table 2, first part. Ellagic acid showed K_i of 16 μM, while apigenin only a poor inhibition of 6% at 100 μM. Considering that apigenin did not fulfill the H-bond potential of TS pockets (Figure 7b) and that flavonoids are known CKII inhibitors, we extended the analyses to other purchased and in-house synthesized flavonoids (compounds 1-4, Table 2, second part; see Materials and Methods for synthesis details). Moreover, both compounds are known to be promiscuous inhibitors and we wanted to further validate the TS inhibition through a second round compounds, similar to the ones selected but with better properties. Docking simulations were used to guide this second selection, with the aim of maintaining the same orientation within the binding site but increasing the number of hydrogen bonds formed with the residues lining the cavity. We ended up with eleven compounds, among which five showed relevant inhibition activity. In particular, seven were purchased from

vendors and four (compounds 1-4) were synthesized in house, as reported in Scheme 1. Among the acquired ones, morin, fisetin, kaempferol presented an inhibition effect in the low micromolar range (K_i equal to 2.9, 3.4, 3.7 μM, respectively). Compounds **1** and **4** showed K_i below 10 μM (K_i of 6.6 and 3.3 μM respectively), compound **2** a poor inhibition of 8% at 25 μM, while **3** could not be studied because of the low solubility. As well as ellagic acid, they all behaved as competitive inhibitors of folic acid, nevertheless most of them presented poor water solubility and we had to test very low inhibitor concentration in the μM tens range.

Scheme 1. A. Synthesis of compounds **1**-**3**. Reaction conditions: a) $SOCl_2$, EtOH, r.t. **B.** Synthesis of compound **4**. Reaction conditions: a) NaOH (3 M), EtOH, r.t. b) H_2O_2,NaOH (1 M), EtOH, r.t. c) BBr_3 (1 M in dry DMC), dry DMC, 0 °C → r.t.



Docking poses of the active compounds within TS binding site are reported in Figure S1. The presence of more polar groups on the molecules allowed the formation of additional hydrogen

bonds within the target pocket, which might justify their higher activity. While apigenin was able to form a few H-bonds (in the pose shown in Figure 7b only the dUMP substrate is contacted) the five active flavonoids interact with Arg50, Glu87, Asn112, Asp218, Val223, Tyr258 and Ala312. This higher protein-ligand structural complementary can increase the inhibitors potency but also reduce the promiscuous character of these molecules. Docking simulations, along with *in vitro* analyses and chemical synthesis, thus demonstrated to be fundamental for compounds selection and optimization.

Both ellagic acid and apigenin are known inhibitors of other proteins apart from CKII. According to Drugbank,[65] ellagic acid is active against carbonic anhydrase 1,2,3,4,5A mitochondrial, 5B mitochondrial, 6, 7, 9, 12, 14, CKIIα, camp-dependent protein kinase alpha, protein kinase alpha and beta type, tyrosine-protein kinase SYK, cytochrome P450 1A1 and 2E1. No specific information is reported for apigenin in Drugbank, while ChEMBL reports a number of possible targets.[66] Among these we find aldehyde dehydrogenase 1A1cytochrome P450 2C9, 2C19 2D6 and 3A4, DNA polymerases, ERBB1, MAP kinase ERK2 and p38 alpha, acetylcholinesterase, cholinesterase, beta secretase 1, piruvate kinase, protein kinase C alpha, Rho-associociated protein kinase 2, tyrosine-protein kinase SYK, phosphodiesterase 5A, and others. Despite their rather promiscuous profile, neither for ellagic acid or apigenin TS was previously reported as a potential target. As well morin, fisetin or kaempferol, according to ChEMBL and PubChem did not count TS, neither human or bacterial, among their possible targets.

Pockets exchange

The above described results suggest that TS and CKIIα pockets present similar properties. We compared the pockets MIFs and reported their superimposition in Figure 8. The pockets shape and dimension present a certain degree of similarity, being the volume 1390 $Å^3$ and 1675 $Å^3$, for TS and CKIIα, respectively (Figure 8a). The hydrophobic MIFs show the higher level of overlap, in

particular in the region where the hydrophobic core of the selected ligands is involved in pi-pi interactions with the dUMP in TS. A lower superimposition can be observed for the H-bond donor and H-bond acceptor group. Clearly, we do not have to expect a total overlap of pockets chemical and geometrical properties, in particular when dealing with such large cavities. Ligands can differently adapt and occupy only portions of the pockets. For instance we have previously reported the similarity of the ERα and the SERCA (sarcoplasmic reticulum Ca^{2+} ion channel ATPase) cavity, which is a known off-target for selective estrogen receptor modulators. The two superposed cavity showed only a 62% overlap of the volume. Again the highest similarity scores were detected for the hydrophobic and the H-bond donor MIFs.[44] Also we have to remember that proteins and ligands are flexible and so is the image that ligands produce in their binding site. We do not expect ligands to be totally complementary towards the new target, but to recognize at least a part of the image they produced in the original pocket.

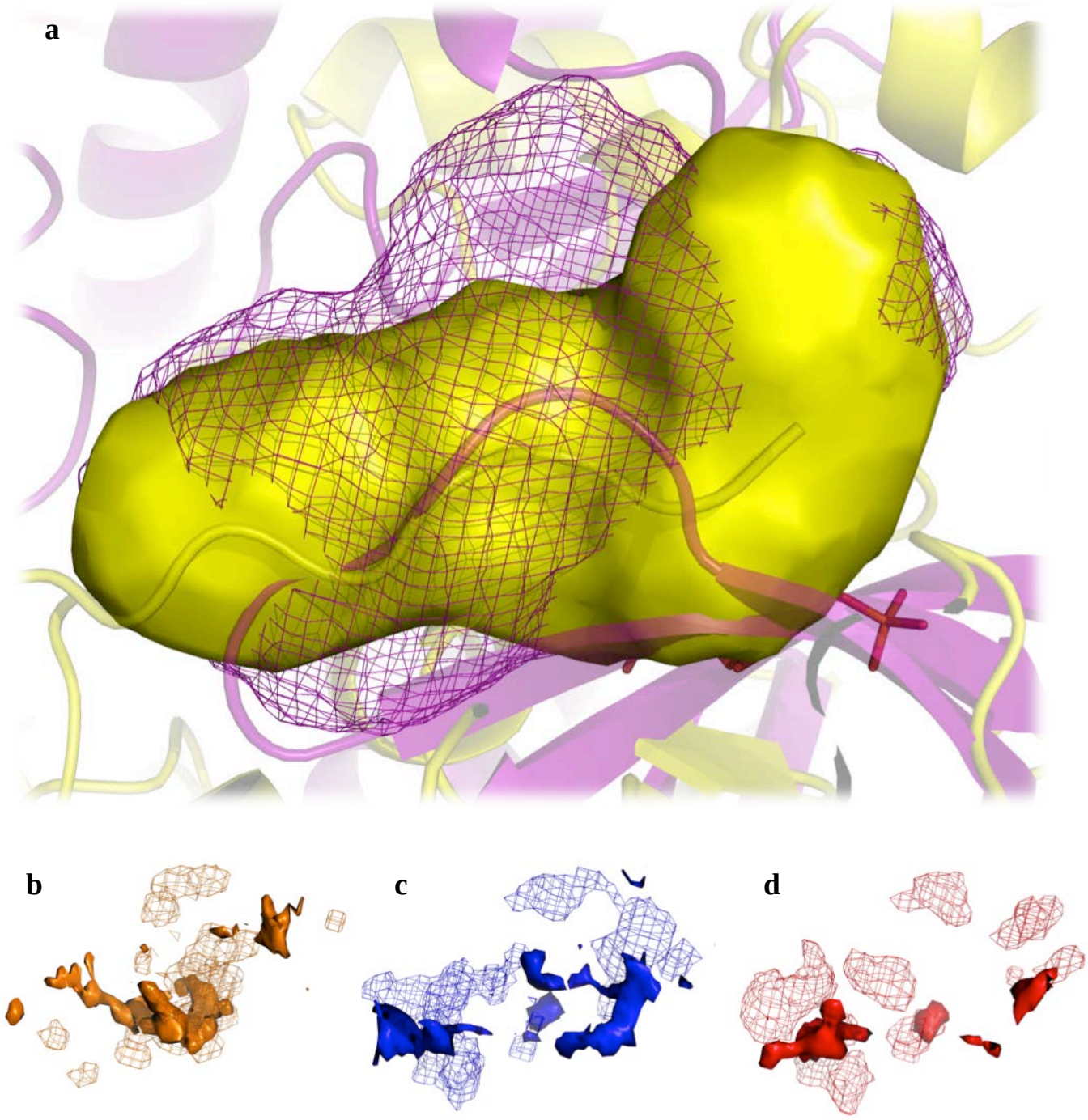


Figure 8. Superimposition of hTS and CK2 binding sites. a. Shape superimposition, TS pocket (PDB code: 1hvy) is shown as yellow surface and CKIIα (PDB code: 2zjw) pocket as magenta mesh lines. **b, c, d.** Superimposition of hydrophobic, H-bond donor and H-bond acceptor molecular interaction fields, respectively; TS MIFs are shown as solid surfaces while CKIIα MIFs as mesh lines

CONCLUSIONS

Recognizing the potential of drug repurposing strategies we presented here the BioGPS/FLAPdock approach for off-targets identification and repositioning issues. Given a protein cavity this approach is able to automatically search the PDB and identify the most similar binding sites in terms of their ligand image, that is, the volume, the shape and the chemical features a ligand encounters once entered a specific pocket. Pockets are represented and compared according to their molecular interaction fields, also encoding the energetics of the pockets, differently from many other approaches. Once similar pockets are identified, co-cristallized ligands, or any other known inhibitor, can be cross-docked between the template and the queries, or among the related queries, looking for new targets and applications.

We run this pipeline for the specific thymidylate synthase case, identifying HSP-90, the estrogen receptor, the vitamin D3 receptor, different kinases, transferarses and phosphodiesterases and others as possible related targets. We selected some ligands from casein kinase II α, ellagic acid and flavonoids, and found them to be inhibitor of our TS template in the low micromolar range.

The specific case here described supports the applicability of the BioGPS/FLAPdock integrated pipeline. Our aim is not to propose ellagic acid or flavonoids as new potential hTS inhibitors, but to point out the possibility of identifying similar proteins in a new, fast and automatic way, and subsequently repurposing known drugs or ligands for specific proteins. As previously described, given a template, the identification of the most similar pockets and, consequently, of the possible off-targets is totally automatic. In the pipeline we reported this corresponds to the first BioGPS step. Only after the selection of possible related targets, ligands can be exchanged by docking simulations, tested and improved by means of *in silico*, *in vitro* analyses, and possible chemical synthesis. This second step allows the rationalization of the protein-ligand interaction, the possible improvement of the complex stability and the identification/development of more potent and specific inhibitors.

The potential of this pipeline is extremely large. Apart from the identification of new inhibitors among known ligands for a specific target, BioGPS has a variety of applications and possible utilities. Pocketomes can be easily and rapidly analyzed for identifying targets likely responsible of unpredicted side-effects. As well, the similarity of targets involved in specific pathways or over-expressed in pathological conditions can be investigated for designing multi-target therapies. The main advantage is represented by the algorithm capability of depicting the real structural and energetic scenario of a protein binding site, totally independent of any other protein or ligand-related information, apart of the pocket itself. As well, the simplicity of the pocket search, the rapid and semi-automated procedure, makes it a promising and valuable tool for polypharmacology, repurposing and side-effects predictions.

MATERIALS AND METHODS

Cavities identification.

The FLAPsite algorithm is used for the identification of cavities in three-dimensional protein structures.[44] By embedding the protein structure into a three-dimensional grid with a spatial resolution of 1.0 Å, the algorithm identifies pocket points using the GRID probe H (shape).[46] For each point a buriedness-index is calculated. Points with a buriedness-index lower than a specific threshold are discarded. Two morphological operations, erosion, and dilation, are applied to the remaining points for removing small anomalies and connecting areas. Hydrophobic probe DRY is used to prioritize hydrophobic cavities usually targeted by drugs. The FLAPsite procedure was applied to (i) human TS (PDBcode 1hvj) with dUMP, without any water molecule (TSp); (ii) human TS (PDBcode 1hvj) with dUMP and two bridging water molecules (wat435 and wat622; TSpw); (iii) all PDB protein structures co-crystallized with a ligand. Only cavities containing a ligand were selected for the following steps of the pipeline: (i-ii) two TS template cavities and (iii) 90.025 dataset cavities.

Virtual screening with BioGPS.

We collected from the Protein Data Bank the structures of all proteins cocrystallized with a ligand. The *fixpdb* tool was used for processing the protein residues, solvent molecules, co-crystallized ligands, cofactors and ions contained in the PDB protein structures. All nucleic acids, ligands and water molecules co-crystallized with the protein were removed, while cofactors were retained (i.e. NAD, FAD, GSH). Additionally, to retain ions involved in interactions with the protein residues, a defined GRID-energy threshold for Cu^{+2}, Fe^{+2}, Zn^{+2}, Mg^{+2} was applied. Binding sites were the detected by using the FLAPsite algorithm (90.025 binding sites; September 2014). MIFs were calculates for each binding site and stored into a database. The BioGPS technology was used to compare two TS pocket templates against 90.025 MIFs cavities dataset. The BioGPS algorithm compares binding sites by means of their Molecular Interaction Fields overlapping,[44,48] and exploits the technology implemented in FLAP.[49] FLAP (Fingerprints for Ligands and Proteins) is a virtual screening algorithm developed and licensed by Molecular Discovery Ltd. (www.moldiscovery.com). Several VS campaigns have been successfully performed with FLAP and are reported in the literature,[67-72] as well as binding modes prediction and rationalization.[73,74] Initially, the approach uses the GRID force field to evaluate the type, strength and direction of the interactions that a cavity is capable of making. The GRID probes H, DRY, O, and N1 are used to compute the shape, the hydrophobic interactions, the H-bond donor interactions and the H-bond acceptor interactions respectively for each cavity considered in the analysis. Since a simple comparison of the entire MIF areas might be computationally expensive, the algorithm reduces the information by selecting a number of representative points, called hotspots, proportionally to the energy and the volume of each MIF. Therefore, all possible combinations of four hotspots (called quadruplets) are generated and stored into a fingerprint, named Common Reference Framework.[49] The BioGPS algorithm compares two cavities by comparing such Common Reference Frameworks in a pair-wise manner. This approach searches

for the largest amount of favourable quadruplets superposition. When the quadruplets of template and candidate cavities match the feature types H (shape), DRY (hydrophobic), O (H-bond donor), N1 (H-bond acceptor) within certain distances (the tolerance is fixed to less than 1 Å), the algorithm overlaps their 3D cavity structures with a specific orientation, according to the quadruplet matching, and then calculate the MIF overlapping areas. The final superposition is called “solution” and is quantitatively scored by considering the corresponding MIFs similarity, collapsed in nineteen different scores. FLAP first calculates scores representing the degree of volume overlap for each of the probes (and of the corresponding generated MIFs) being used individually, i.e. H, DRY, O and N1, and then combines these scores in order to produce probe-combination scores. In addition FLAP calculates two Global scores, the Global Sum, which is produced by summing all the scores of the individual probes together, and the Global Product (GlobP), produced by multiplying all the scores of the individual probes together. Once the Probe scores for the individual probes and their combinations have been calculated, including the Global Sum and Global Product, FLAP will calculate a Distance Score, representing the overall similarity derived by a combination of the overlap degree between the single H, DRY, O and N1 MIFs computed for the candidates and the template, that is, the protein binding site. The Global Product score was considered for hits selection in this study; it spans between 0 and 1, where the higher the score the more similar are the two entities.

Hits (pockets) selection.

The Global Product score was set to 0.7 as restrictive threshold for selecting the most similar cavities to the hTS templates; 4.476 cavities were selected. Then, for each cavity the 'fraction of ligand volume’ (FV) was calculated as the ratio of the ligand volume contained within a cavity over the total volume of the ligand. Only cavities with FV greater than 0.5 were selected. We thus obtained 3.770 cavities, from which 1.361 co-crystallized ligands were extracted. Non relevant

compounds as solvents, i.e. ethylene glycol, glycerol, or prosthetic groups were discarded. The remaining ligands were compared to a set of known human TS inhibitors, retrieved from the ChEMBL database.[66] A bioactivity search was performed and 354 inhibitors with known K_i towards human TS were retrieved. Molecules likely designed to displace the dUMP from the binding site were discarded. The remaining 328 molecules were analyzed with Volsurf to calculate their volume. The minimum and maximum volumes were used as cut-offs. Unknown ligands (extracted form cavities) having a volume higher than the smallest hTS inhibitor and lower than the biggest one were retained. We ended up with 283 ligands belonging to 135 proteins and 317 pockets (Table S1).

Connectivity analysis.

Network graphs reported in Figure 5 and 6b were built with Cytoscape.[59] In the BioGPS network (Figure 5) the GlobP score was used to define the distance among the 317 pockets. To calculate a GlobP value for each pocket pair, a BioGPS "all *vs* all" VS approach was adopted and each cavity was compared and scored towards the others. Since the distance is meant to be lower than a threshold value to connect two objects in a network graph, pocket pairs were joined when having a 1 - GlobP value lower than 0.6, that is GlobP > 0.4. To avoid over-connections between TS and the other cavities in the graph (all cavities identified as similar from the first BioGPS VS had at least GlobP > 0.7 with respect to TS), more strict threshold were set. 1 - GlobP < 0.2 (GlobP > 0.8) and 1 - GlobP < 0.05 (GlobP > 0.95) were the thresholds considered to connect TS with the other pockets and to itself, respectively.

In the Volsurf network (Figure 6b), distances were taken from the square matrix obtained calculating the distance between each pair of objects in the PCA bidimensional space. Pockets were connected when having a distance < 0.8. Pockets sharing no similarity with any other element in the set were classified as singletons.

Volsurf-based pocket analysis.

Volsurf-type descriptors are based on Volsurf software,[61] developed to calculate pharmacokinetic properties from 3D Molecular Interaction Fields.[72,75] Each protein cavity is described by the corresponding MIFs, and no comparison algorithm is used. The aim of the Volsurf-type approach is to separate cavities accordingly to their morphological properties and to the principal interactions they might undertake with a putative ligand. After each active site is detected and MIFs are calculated by using default GRID probes (H, O, N1, DRY), the second step consists in the calculation of descriptors from the 3D maps obtained. The molecular descriptors obtained, refer to protein cavity size and shape, and to the volume of hydrophilic and hydrophobic regions. The VolSurf-type descriptors are mainly of two types: morphological descriptors (volume, surface, rugosity, globularity) and energy-dependent descriptors (volume of H-bond and hydrophobic interactions at different energy levels). The descriptors obtained can be analyzed by Principal Component Analysis (PCA) to reduce the dimensionality of the data and to separate cavities accordingly to their physical-chemical properties. PCA converts a set of correlated descriptors into a new set of linearly independent variables. Each component is calculated in order to maximize the variance of the object in the dataset.

Considering components scores (e.g. PC1 vs PC2) it is possible to project the object in a bi-dimensional space (e.g. x axes reports score values calculated on PC1, y axes reports score values calculated on PC2), where the objects falling close to each other present similar properties (similar VS descriptors, similar MIFs volumes).

A square matrix is obtained calculating the distance in the bidimensional space between each pair of objects. This matrix can be used for a network-based analysis in to easily inspect and analyse the similarities and differences among the set of objects (cavities).

Molecular Docking.

FLAPdock is a docking approach implemented in the software FLAP,[49] based on GRID Molecular Interaction Field similarities,[46] combined with classical energetics. FLAPdock follows a molecular fragmentation approach, subsequent placement of each fragment in the site of the target, followed by incremental construction of the molecule. At each phase of the docking, a number of solutions are generated, scored, and a subset retained for the subsequent phase. A set of poses for the starting fragment is generated using FLAP quadruplet alignment of the fragment conformer atom quadruplets, and the receptor site GRID MIF minima points. In this way hundreds of thousands of poses are typically generated for this starting fragment. As a first step, the poses are scored using a weighted sum of the FLAP field similarities, including shape, donor, acceptor, and hydrophobic similarity. A second scoring step calculates Lennard-Jones and dielectric corrected Coulombic energetic terms for this subset; the solutions are then ranked according to the combined score and, subjected to RMS clustering, and the best scoring pose in each cluster retained.

Internal validation (re-docking of x-ray ligands) has shown that one of the top five poses contains a pose within 2.0 Å of the x-ray position in more than 90% of the cases [unpublished results]. Recent simulations of covalent docking demonstrated to provide reliable results in agreement with experimental data [work in preparation]. The most promising compounds were finally selected according to the FLAP S score value, to the number of hydrogen bonds made with the surrounding residues and to the complementarity of the pocket MIFs with the ligand pseudoMIFs. PseudoMIFs correspond to the projection of the MIFs on the atoms that generate them.

Ligands tautomers and protomers enumeration was performed with Moka 2.6 before docking simulations.

Docking analyses were performed within both TSp and TSpw.

Protein expression and purification.

Human thymidylate synthase was cloned in the pQE80L system, as reported in [76]. The recombinant protein was expressed in DH5a *Escherichia coli* strain. The expression vector codes for a hexa-histidine tag at the N-terminus of the gene product, designed to facilitate the purification of the recombinant protein thorough immobilized metal affinity chromatography.

Bacteria (DH5a/pQE80L) were grown in 2 L of LB medium containing 50 mg ml^{-1} ampicillin. The solution was centrifuged at 37°C, 120 rpm, until the OD_{600nm} reached a value of 0.6. TS expression was induced adding 1 mM isopropyl-b-D-thiogalactopyranoside (IPTG) and incubating the culture for 4 hours, 37°C, 120 rpm. Cells were then centrifuged at 4000 rpm for 30 min at 4°C. The cell pellet was suspended in buffer A (20 mM NaH_2PO_4, 30 mM NaCl, 20 mM Imidazol, pH 7.5) containing Complete® (protease inhibitor) and sonicated in cooling ice bath. The broken cells were centrifuged for 40 minutes at 12000 rpm, 4°C, and the pellet was discarded. The supernatant was treated with streptomycin 10%, stirred for 10 minutes at 4°C and centrifuged for 30 minutes, 12000 rpm, 4°C. The discarded pellet and the supernatant were filtered by 0,8/0,45 µm filters and loaded on a Ni-HTP column preequilibrated with buffer A. The enzyme was eluted with buffer B (20 mM NaH_2PO_4, 30 mM NaCl, 1 M Imidazol, pH 7.5). The fractions with enzyme were collected in pool and loaded on a column HiTrap Desalting to change the buffer with buffer C (20 mM NaH_2PO_4, 30 mM NaCl, pH 7,5).[77] Only fractions with detected enzymatic activity were collected.

Enzymatic activity and inhibition assays.

TS enzymatic activity was measured spectrophotometrically (Beckman DU640) by monitoring the absorbance increase at 340 nm, for 3 minutes during the oxidation reaction of the substrate THF to 7,8-dihydrofolate. K_M values were determined for both mTHF and dUMP varying the substrate concentrations. The concentration ranges for K_M were 2–80 µM for mTHF and 3-150 µM for

dUMP. Values of k_{cat} and specific activity were determined by varying the enzyme concentration in a 0,04–0,3 µM range. The reaction mixture contained 50% of assay buffer (TES, N-[tris(hydroxymethyl)methyl]-2-aminoethanesulphonic acid 100 mM, MgCl2 50 mM, formalin 13 mM, EDTA 2 mM, pH 7.4 b-mercaptoethanol 150 mM), the enzyme at a concentration of 0.1 µM, 50 µM of mTHF, 120 µM of dUMP and water to 800 µl. The reaction was initiated when dUMP was added to the reaction mixture.

The selected compounds were evaluated against recombinant hTS and the inhibition percentage was determined for a 10-100 µM compound concentration range. The molecules were solubilized in DMSO at 10 mM concentration.[77] The inhibition percentage was determined upon evaluation of the DOD/min ratio.

It was not possible to perform a detailed study of the inhibition activity for all the compounds because of their poor aqueous solubility. Different samples at 100, 50, 25 or 10 µM concentration were prepared, reducing the concentration until opalescence in the solution and/or scattering effects disappeared. Thus, inhibition assays were performed at the compounds maximum solubility. Given the impossibility to gradually increase the compounds concentration and experimentally determine the IC_{50} value, IC_{50} and K_i were calculated from the inhibition percentage as reported in.[78] Values might be underestimated.

Synthetic chemistry.

The following compounds were purchased from Sigma-Aldrich: ellagic acid (CAS: 476-66-4), apigenin (CAS: 520-36-5), morin hydrate (CAS: 654055-01-3), fisetin (CAS: 345909-34-4), datiscetin (CAS: 480-15-9), taxifolin (CAS: 480-18-2), (+)-catechin (CAS: 154-23-4), kaempferol (CAS: 520-18-3), quercetin (CAS: 117-39-5). In Table 2 the molecular weight of morin, fisetin, quercetin is reported as anhydrous basis. All the reagents and solvents used for the synthesis of compounds **1-4** were purchased from Sigma-Aldrich and used without further purification. Silica

gel plates (Merck F254) were used for thin-layer chromatography. ^{1}H and ^{13}C NMR spectra were recorded on a Bruker FT-NMR AVANCE 400. Chemical shifts (δ scale) are reported in parts per million downfield from tetramethylsilane as internal standard. Splitting patterns are designated as follows: s, singlet; d, doublet; t, triplet; q, quadruplet; m, multiplet; br s, broad singlet; and dd, double doublet. Silica gel Merck (60-230 mesh) was used for column chromatography. Melting points were determined with a Stuart SMP3 and they are uncorrected. Mass spectra were obtained on a 6520 Accurate-Mass Q-TOF LC/MS.

The synthetic procedures for the synthesis of compounds **1-4** are reported in Scheme 1A and 1B. Compounds **5** and **6** are the intermediates for the synthesis of compound **4**.

General procedure for the synthesis of hydroxylated flavanones (1, 2 and 3). To a stirred mixture of 2',5'-dihydroxyacetophenone (0.300 g, 1.97 mmol) and the appropriate aldehydes (1 eq.) in absolute ethanol (2ml), thionyl chloride (120 μL) was added dropwise over 5 minutes. The reaction was stirred at room temperature for 6 hours. Ethanol and excess of thionyl chloride were removed under reduced pressure on a rotary evaporator. Column chromatography was carried out to purify the desired product (eluent system: cyclohexane/ethyl acetate 9.8/0.2).

6-hydroxy-2-(4-hydroxyphenyl)chroman-4-one (1) was isolated as a yellow solid in a 31% yield. Mp 230 °C. ^{1}H NMR (CD_3OD, 400 MHz) δ ppm: 7.33 (d, 2H, $J_{2',3'/5',6'}$ = 8.6 Hz, H-2' + H-6'), 7.22 (d, 1H, $J_{5,7}$ = 3.0 Hz, H-5), 7.04 (dd, 1H, $J_{7,8}$ = 8.9 Hz, $J_{7,5}$ = 3.0 Hz, H-7), 6.90 (d, 1H, $J_{8,7}$ = 8.9 Hz, H-8), 6.84 (d, 2H, $J_{3',2'/5',6'}$ = 8.6 Hz, H-3' + H-5'), 5.34 (dd, 1H, $J_{2,3b}$ = 13.3 Hz, $J_{2,3a}$ = 2.8 Hz, H-2), 3.08 (dd, 1H, $J_{3b,3a}$ = 17.0 Hz, $J_{3b,2}$ = 13.3 Hz, Hb-3), 2.74 (dd, 1H, $J_{3a,3b}$ = 17.0 Hz, $J_{3a,2}$ = 2.8 Hz, Ha-3). ^{13}C NMR (CD_3OD, 100 MHz) δ ppm: 193.46, 157.50, 155.63, 151.51, 130.09, 127.59 (2C), 124.55, 120.77, 118.76, 114.90 (2C), 109.95, 79.45, 44.00. ESI-HRMS calcd for $C_{15}H_{13}O_4$ $[M+H]^+$ 257.0808, found 257.0805.

6-hydroxy-2-(3-hydroxyphenyl)chroman-4-one (2) was isolated as a yellow solid in a 26% yield. Mp 240 °C. ^{1}H NMR (DMSO, 400 MHz) δ ppm: 9.51 (br s, 1H, 3'-OH), 9.42 (br s, 1H, 6-OH), 7.20 (dd, 1H, $J_{5',4'}$ = 8.2 Hz, $J_{5',6'}$ = 7.4 Hz, H-5'), 7.12 (d, 1H, $J_{5,7}$ = 2.9 Hz, H-5), 7.04 (dd, 1H, $J_{7,8}$ = 8.8 Hz, $J_{7,5}$ = 2.9 Hz, H-7), 6.95 (d, 1H, $J_{8,7}$ = 8.8 Hz, H-8), 6.91 (m, 2H, H-6' + H-2'), 6.76 (d, 1H, $J_{4',5'}$ = 8.2 Hz, H-4'), 5.47 (dd, 1H, $J_{2,3b}$ = 12.7 Hz, $J_{2,3a}$ = 2.6 Hz, H-2), 3.10 (dd, 1H, $J_{3b,3a}$ = 16.8 Hz, $J_{3b,2}$ = 12.7 Hz, Hb-3), 2.76 (dd, 1H, $J_{3a,3b}$ = 16.8 Hz, $J_{3a,2}$ = 2.6 Hz, Ha-3). ^{13}C NMR (DMSO, 100 MHz) δ ppm: 192.17, 157.90, 154.81, 152.03, 141.11, 129.99, 124.98, 121.33, 119.46, 117.43, 115.69, 113.79, 110.39, 79.08, 44.22. ESI-HRMS calcd for $C_{15}H_{13}O_4$ $[M+H]^+$ 257.0808, found 257.0808.

2-(3,4-dihydroxyphenyl)-6-hydroxychroman-4-one (3) was isolated as an orange solid in a 26% yield. Mp 220 °C. ^{1}H NMR (DMSO, 400 MHz) δ ppm: 9.35 (br s, 1H, OH), 8.90 (br s, 2H, OH), 7.11 (d, 1H, $J_{5,7}$ = 3.0 Hz, H-5), 7.02 (dd, 1H, $J_{7,8}$ = 8.8 Hz, $J_{7,5}$ = 3.0 Hz, H-7), 6.91 (d, 1H, $J_{8,7}$ = 8.8 Hz, H-8), 6.90 (m, 1H, H-2'), 6.75 (m, 2H, H-5' + H-6'), 5.35 (dd, 1H, $J_{2,3b}$ = 12.8 Hz, $J_{2,3a}$ = 2.6 Hz, H-2), 3.10 (dd, 1H, $J_{3b,3a}$ = 16.8 Hz, $J_{3b,2}$ = 12.8 Hz, Hb-3), 2.68 (dd, 1H, $J_{3a,3b}$ = 16.8 Hz, $J_{3a,2}$ = 2.6 Hz, Ha-3). ^{13}C NMR (DMSO, 100 MHz) δ ppm: 192.52, 154.99, 151.90, 146.01, 145.62, 130.49, 124.93, 121.26, 119.42, 118.27, 115.77, 114.74, 110.36, 79.22, 44.13. ESI-HRMS calcd for $C_{15}H_{13}O_5$ $[M+H]^+$ 273.0757, found 273.0759.

Synthesis of (E)-3-(3,4-dimethoxyphenyl)-1-(2-hydroxy-5-methoxyphenyl)prop-2-en-1-one (6). To a solution of 2'-hydroxy-5'-methoxyacetophenone (0.724 g, 4.36 mmol) and 3,4-dimethoxybenzaldehyde (0.724 g, 4.36 mmol) in EtOH (12 mL), an aqueous solution of NaOH (3 M, 4.4 mL) was added. The reaction was stirred at room temperature overnight. The reaction mixture was cooled in an ice-water bath and acidified to pH 2 with concentrated HCl (37%). The solid formed was filtered and re-crystallized from ethanol to obtain compound **6** as a red solid (0.781 g, 57% yield). Mp 114-115 °C. ^{1}H NMR ($CDCl_3$, 400 MHz) δ ppm: 12.29 (br s,1H, OH), 7.92 (d, 1H, $J_{B,A}$= 15.7 Hz, H_B), 7.49 (d, 1H, $J_{A,B}$ = 15.7 Hz, H_A), 7.43 (d, 1H, $J_{6',4'}$ = 2.9 Hz, H-

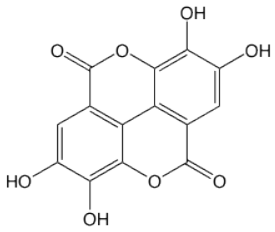
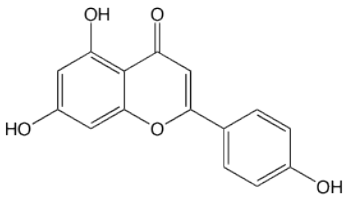
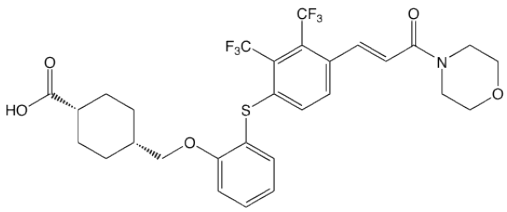
6'), 7.31 (dd, 1H, $J_{6,5}$= 8.3, $J_{6,2}$ = 2.0 Hz, H-6), 7.20 (d, 1H, $J_{2,6}$ = 2.0 Hz, H-2), 7.18 (dd, 1H, $J_{4',3'}$ = 9.3, $J_{4',6'}$ = 2.9 Hz, H-4'), 7.01 (d, $J_{3',4'}$ = 9.3 Hz, H-3'), 6.95 (d, 1H, $J_{5,6}$ = 8.3 Hz, H-5), 4.01 (s, 3H, -OCH3), 3.98 (s, 3H, -OCH3), 3.88 (s, 3H, -OCH3). ^{13}C NMR ($CDCl_3$, 100 MHz) δ ppm: 193.27, 157.84, 151.90, 151.65, 149.36, 145.87, 127.61, 123.61, 123.17, 119.87, 119.25, 117.83, 113.54, 111.22, 110.50, 101.01, 56.25, 56.09, 56.07. ESI-HRMS calcd for $C_{18}H_{19}O_5$ $[M+H]^+$ 315.1227, found 315.1232.

Synthesis of 2-(3,4-dimethoxyphenyl)-3-hydroxy-6-methoxy-4H-chromen-4-one (5). An aqueous solution of H_2O_2 (30%, 620 µL) was added to an ice-cold suspension of compound **6** (0.781 g, 2.48 mmol) in ethanol (12 mL) and 1 M NaOH (5 mL). The mixture was allowed to warm to room temperature and was stirred for 4 hours. Then the reaction mixture was cooled in an ice-water bath and acidified to pH 2 with concentrated HCl (37%). The solid formed was filtered and re-crystallized from ethanol to obtain compound **5** as pale-yellow powder (0.522 g, 64% yield). Mp 190 °C. ^{1}H NMR ($CDCl_3$, 400 MHz) δ ppm: 7.88 (dd, $J_{6',2'}$ = 1.9 Hz, $J_{6',5'}$ = 8.6 Hz, H-6'),7.84 (s, 1H, $J_{2',6'}$ = 1.9 Hz, H-2'),7.55 (d, 1H, $J_{5,7}$ = 3.0 Hz, H-5), 7.51 (d, 1H, $J_{8,7}$ = 9.1 Hz, H-8), 7.30 (dd, 1H, $J_{7,8}$ = 9.1 Hz, $J_{7,5}$ = 3.0 Hz, H-7), 7.14 (br s, 1H, OH), 7.01 (d, 1H, $J_{5',6'}$ = 8.6 Hz, H-5'), 3.98 (s, 3H, -OCH3), 3.93 (s, 3H, -OCH3), 3.88 (s, 3H, -OCH3). ^{13}C NMR ($CDCl_3$, 100 MHz) δ ppm: 172.70, 156.42, 150.69, 150.28, 148.82, 145.07, 137.51, 124.09, 123.79, 121.42, 121.11, 119.58, 110.91, 110.68, 103.83, 56.01, 55.96, 55.92. ESI-HRMS calcd for $C_{18}H_{17}O_6$ $[M+H]^+$ 329.1020, found 329.1024.

Synthesis of 2-(3,4-dihydroxyphenyl)-3,6-dihydroxy-4H-chromen-4-one (4). To a stirring solution of compound **5** (0.500 g, 1.52 mmol) in anhydrous dichloromethane (30 mL) under nitrogen at 0°C, boron tribromide in dichloromethane (1.0 M, 13.7 mL, 13.7 mmol) was added. The mixture was allowed to warm to room temperature and stirred for 2 days. The reaction mixture was then cooled to 0°C and methanol (10 mL) was added. The reaction mixture was concentrated *in vacuo*. Water (10mL) was added, the reaction was sonicated for 1 hour and then

left to stand. The solid was filtered to collect compound **4** as a red solid (0.440 g, quantitative yield). Mp 336 °C dec. ^{1}H NMR (DMSO, 400 MHz) δ ppm: 7.73 (d, 1H, $J_{8,7}$ = 2.2 Hz, H-2'), 7.58 (dd, 1H J = 2.2, J = 8.5 Hz, H-6'), 7.57 (d, 1H, J = 9.1 Hz, H-8), 7.35 (d, 1H, J = 3.0 Hz, H-5), 7.23 (dd, 1H, $J_{7,8}$ = 9.1 Hz, $J_{7,5}$ = 3.0 Hz, H-7), 6.90 (d, 1H, J = 8.5 Hz, H-5'). ^{13}C NMR (DMSO, 100 MHz) δ ppm: 172.58, 154.46, 148.82, 147.92, 146.30, 145.50, 137.72, 123.35, 122.95, 122.52, 120.36, 119.95, 116.03, 115.66, 107.27. ESI-HRMS calcd for $C_{15}H_{11}O_6$ $[M+H]^+$ 287.0550, found 287.0552.

Table 1. Chemical structure and properties of the most promising compounds selected by the BioGPS-FLAPdock integrated approach.

Compound PDB code	Chemical structure	Co-crystallized protein (PDB code)
REF		Casein kinase II α (2zjw)
AGI		Casein kinase II α (3amy)
E2M		LFA-1 binding domain (3e2m)

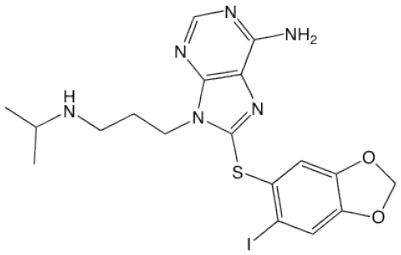
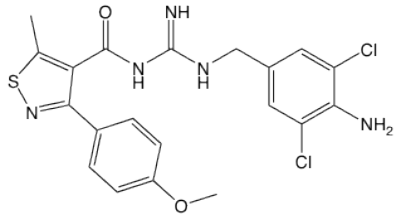
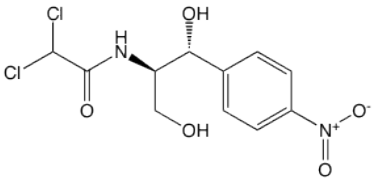
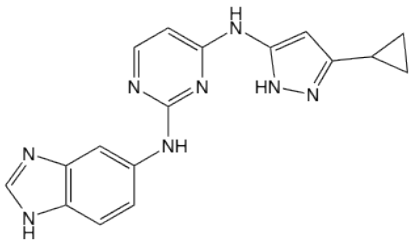
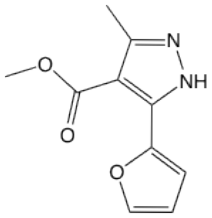
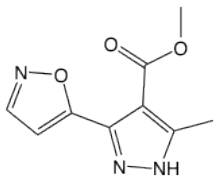
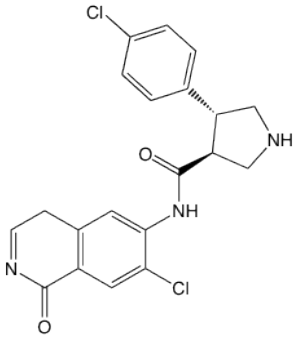
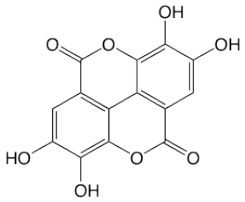
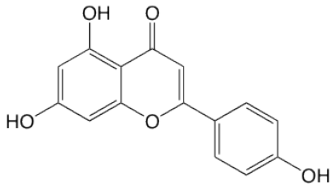
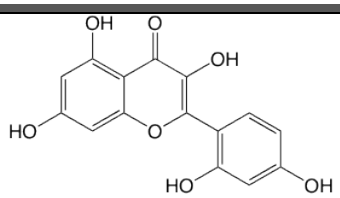
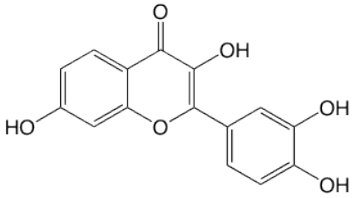
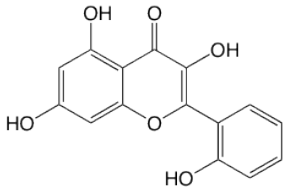
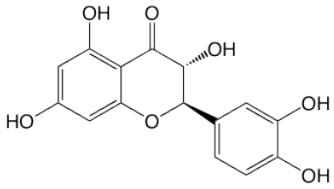
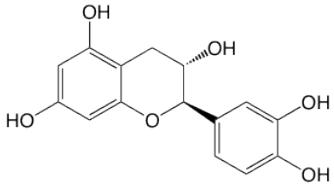
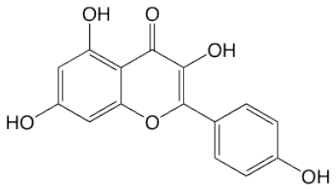
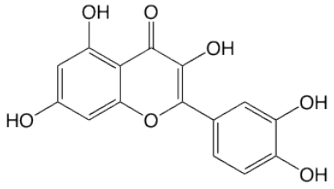
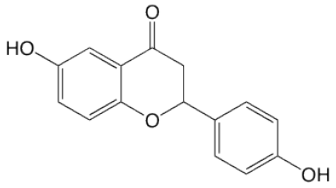
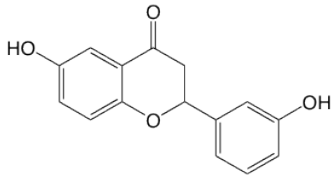
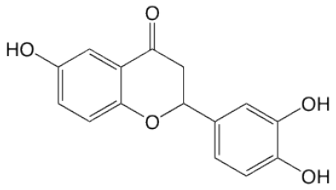
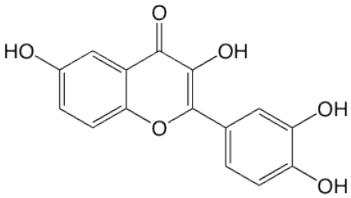
H71		HSP-90 (2fwz)
0VA		Beta-secretase 1 (4fse)
CLM		Chloramphenicol acetyltransferase 3 (4cla)
APJ		Serine/threonine-protein kinase/ endoribonuclease IRE1 (3fbv)
37D		HSP-90 (3hzi)
2D3		HSP-90 (2ye8)
3ND		Rho associated protein kinase (3ndm)

Table 2. Chemical structures and inhibition values for the compounds tested against hTS.

Compound name	MW	Compound structure	[a]IC50/Ki (μM)
Ellagic acid	302.19		238/16
Apigenin	270.24		1440/94
Morin	302.24		45/2.9
Fisetin	286.24		52/3.4
Datiscetin	286.24		[b]NI
Taxifolin	304.25		NI
Catechin	290.27		NI

Kaempferol	286.24		57/3.7
Quercetin	302.24		NI
Compound 1	256.25		101/6.6
Compound 2	256.25		NI
Compound 3	272.25		273/18
Compound 4	286.24		61/4.0

[a] Standard error within +/- 20% of the given value.[54]
[b] NI = no inhibition at the solubility limit.

SUPPORTING INFORMATION

List of 317 selected cavities, along with their PDB, UNIPROT codes and co-crystallized ligands. Docking poses of active flavonoids compounds within TSpw. Loading plots of the physicochemical properties distribution for the 317 cavities in the multivariate space.

ACKNOWLEDGMENTS

We kindly acknowledge Molecular Discovery Ltd for supplying the BioGPS the FLAPdock softwares, for the virtual screening and docking analyses.

We also acknowledge AIRC 2015, grant IG10474 and IG16977 given to MPC.

ABBREVIATIONS

CKIIα, casein kinase II alpha; dUMP, deoxyuridine monophosphate; DOD, differential optical depth; ER, estrogen receptor; FV, fraction of ligand volume; GlobP, Global Product; IPTG, isopropyl-b-D-thiogalactopyranoside; MIFs, Molecular Interaction Fields; PCA, Principal Component Analysis; PDE, phosphodiesterase; SEs, side effects; THF, tetrahydrofolate; TS, human thymidylate synthase; TSp, human thymidylate synthase pocket; TSpw, human thymidylate synthase pocket with waters; VS, virtual screening.

REFERENCES

[1] D. Farina, F. Spyrakis, A. Venturelli, S. Cross, D. Tondi, M. P. Costi, *Curr. Med. Chem.* **2014**, *21*, 1405-34.
[2] F. Spyrakis, *Curr. Drug Targets* **2015**.
[3] D. Tondi, S. Cross, A. Venturelli, M. P. Costi, G. Cruciani, F. Spyrakis, *Curr. Drug Targets* **2015**.
[4] C. Holohan, S. Van Schaeybroeck, D. B. Longley, P. G. Johnston, *Nat. Rev. Cancer* **2013**, *13*, 714-26.
[5] B. Witkop, *Proc. Am. Philos. Soc.* **1999**, *143*, 540-57.
[6] A. L. Hopkins, *Nature* **2009**, *462*, 167-8.
[7] G. V. Paolini, R. H. Shapland, W. P. van Hoorn, J. S. Mason, A. L. Hopkins, *Nat. Biotechnol.* **2006**, *24*, 805-15.
[8] M. A. Yildirim, K. I. Goh, M. E. Cusick, A. L. Barabasi, M. Vidal, *Nat. Biotechnol.* **2007**, *25*, 1119-26.
[9] A. L. Hopkins, *Nat. Chem. Biol.* **2008**, *4*, 682-90.
[10] T. I. Oprea, J. Mestres, *AAPS J.* **2012**, *14*, 759-63.
[11] Z. Simon, A. Peragovics, M. Vigh-Smeller, G. Csukly, L. Tombor, Z. Yang, G. Zahoranszky-Kohalmi, L. Vegner, B. Jelinek, P. Hari, C. Hetenyi, I. Bitter, P. Czobor, A. Malnasi-Csizmadia, *J. Chem. Inf. Model.* **2012**, *52*, 134-45.
[12] A. S. Reddy, S. Zhang, *Expert Rev. Clin. Pharmacol.* **2013**, *6*, 41-7.
[13] J. Bowes, A. J. Brown, J. Hamon, W. Jarolimek, A. Sridhar, G. Waldron, S. Whitebread, *Nat. Rev. Drug Discov.* **2012**, *11*, 909-22.
[14] T. T. Ashburn, K. B. Thor, *Nat. Rev. Drug. Discov.* **2004**, *3*, 673-83.
[15] C. R. Chong, D. J. Sullivan, Jr., *Nature* **2007**, *448*, 645-6.
[16] A. Bairoch, *Nucleic Acids Res.* **1991**, *19 Suppl*, 2241-5.
[17] A. Bairoch, B. Boeckmann, *Nucleic Acids Res.* **1994**, *22*, 3578-80.
[18] A. J. Bleasby, D. Akrigg, T. K. Attwood, *Nucleic Acids Res.* **1994**, *22*, 3574-7.
[19] W. R. Pearson, *Methods Enzymol.* **1990**, *183*, 63-98.
[20] O. Roche, P. Schneider, J. Zuegge, W. Guba, M. Kansy, A. Alanine, K. Bleicher, F. Danel, E. M. Gutknecht, M. Rogers-Evans, W. Neidhart, H. Stalder, M. Dillon, E. Sjogren, N. Fotouhi, P. Gillespie, R. Goodnow, W. Harris, P. Jones, M. Taniguchi, S. Tsujii, W. von der Saal, G. Zimmermann, G. Schneider, *J. Med. Chem.* **2002**, *45*, 137-42.
[21] N. Atias, R. Sharan, *J. Comput. Biol.* **2011**, *18*, 207-18.
[22] S. Mizutani, E. Pauwels, V. Stoven, S. Goto, Y. Yamanishi, *Bioinformatics* **2012**, *28*, i522-i528.
[23] E. Pauwels, V. Stoven, Y. Yamanishi, *BMC Bioinformatics* **2011**, *12*, 169.
[24] Y. Yamanishi, M. Kotera, M. Kanehisa, S. Goto, *Bioinformatics* **2010**, *26*, i246-54.
[25] Y. Yamanishi, M. Kotera, Y. Moriya, R. Sawada, M. Kanehisa, S. Goto, *Nucleic Acids Res.* **2014**, *42*, W39-45.
[26] V. I. Perez-Nueno, M. Souchet, A. S. Karaboga, D. W. Ritchie, *J. Chem. Inf. Model.* **2015**, *55*, 1804-23.
[27] M. Campillos, M. Kuhn, A. C. Gavin, L. J. Jensen, P. Bork, *Science* **2008**, *321*, 263-6.
[28] J. Scheiber, B. Chen, M. Milik, S. C. Sukuru, A. Bender, D. Mikhailov, S. Whitebread, J. Hamon, K. Azzaoui, L. Urban, M. Glick, J. W. Davies, J. L. Jenkins, *J. Chem. Inf. Model.* **2009**, *49*, 308-17.
[29] L. Yang, P. Agarwal, *PLoS One* **2011**, *6*, e28025.
[30] F. Cheng, W. Li, Z. Wu, X. Wang, C. Zhang, J. Li, G. Liu, Y. Tang, *J. Chem. Inf. Model.* **2013**, *53*, 753-62.
[31] R. B. Altman, *Nat. Genet.* **2007**, *39*, 426.
[32] M. Kanehisa, S. Goto, M. Furumichi, M. Tanabe, M. Hirakawa, *Nucleic Acids Res.* **2010**, *38*, D355-60.
[33] M. Kuhn, M. Campillos, I. Letunic, L. J. Jensen, P. Bork, *Mol. Syst. Biol.* **2010**, *6*, 343.
[34] M. Duran-Frigola, P. Aloy, *Chem. Biol.* **2013**, *20*, 594-603.
[35] V. J. Haupt, S. Daminelli, M. Schroeder, *PLoS One* **2013**, *8*, e65894.

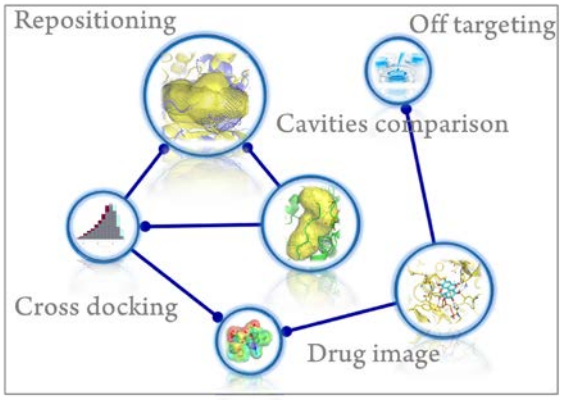
[36] F. Moriaud, S. B. Richard, S. A. Adcock, L. Chanas-Martin, J. S. Surgand, M. Ben Jelloul, F. Delfaud, *Brief Bioinform.* **2011**, *12*, 336-40.
[37] J. Bostrom, A. Hogner, S. Schmitt, *J. Med. Chem.* **2006**, *49*, 6716-25.
[38] Y. Hu, J. Bajorath, *Drug Discov. Today* **2013**, *18*, 644-50.
[39] K. Park, S. Lee, H. S. Ahn, D. Kim, *Mol. Biosyst.* **2009**, *5*, 844-53.
[40] S. Schmitt, D. Kuhn, G. Klebe, *J. Mol. Biol.* **2002**, *323*, 387-406.
[41] A. Shulman-Peleg, R. Nussinov, H. J. Wolfson, *J. Mol. Biol.* **2004**, *339*, 607-33.
[42] N. Weill, D. Rognan, *J. Chem. Inf. Model.* **2010**, *50*, 123-35.
[43] K. Yeturu, N. Chandra, *BMC Bioinformatics* **2008**, *9*, 543.
[44] L. Siragusa, S. Cross, M. Baroni, L. Goracci, G. Cruciani, *Proteins* **2015**, *83*, 517-32.
[45] L. Siragusa, F. Spyrakis, L. Goracci, S. Cross, G. Cruciani, *Molecular Informatics* **2014**, *33*, 446-453.
[46] P. J. Goodford, *J. Med. Chem.* **1985**, *28*, 849-57.
[47] E. Carosati, S. Sciabola, G. Cruciani, *J Med Chem* **2004**, *47*, 5114-25.
[48] V. Ferrario, L. Siragusa, C. Ebert, M. Baroni, M. Foscato, G. Cruciani, L. Gardossi, *PLoS One* **2014**, *9*, e109354.
[49] M. Baroni, G. Cruciani, S. Sciabola, F. Perruccio, J. S. Mason, *J. Chem. Inf. Model.* **2007**, *47*, 279-94.
[50] C. W. Carreras, D. V. Santi, *Annu. Rev. Biochem.* **1995**, *64*, 721-62.
[51] E. Chu, M. A. Callender, M. P. Farrell, J. C. Schmitz, *Cancer Chemother. Pharmacol.* **2003**, *52 Suppl 1*, S80-9.
[52] K. Azijli, I. A. van Roosmalen, J. Smit, S. Pillai, M. Fukushima, S. de Jong, G. J. Peters, I. V. Bijnsdorp, F. A. Kruyt, *Cancer Chemother. Pharmacol.* **2014**, *73*, 1273-83.
[53] E. Carosati, A. Tochowicz, G. Marverti, G. Guaitoli, P. Benedetti, S. Ferrari, R. M. Stroud, J. Finer-Moore, R. Luciani, D. Farina, G. Cruciani, M. P. Costi, *J. Med. Chem.* **2012**, *55*, 10272-6.
[54] M. P. Costi, A. Gelain, D. Barlocco, S. Ghelli, F. Soragni, F. Reniero, T. Rossi, A. Ruberto, C. Guillou, A. Cavazzuti, C. Casolari, S. Ferrari, *J. Med. Chem.* **2006**, *49*, 5958-68.
[55] D. Tondi, A. Venturelli, S. Ferrari, S. Ghelli, M. P. Costi, *J. Med. Chem.* **2005**, *48*, 913-6.
[56] J. Phan, S. Koli, W. Minor, R. B. Dunlap, S. H. Berger, L. Lebioda, *Biochemistry* **2001**, *40*, 1897-902.
[57] G. Ausiello, P. F. Gherardini, E. Gatti, O. Incani, M. Helmer-Citterich, *BMC Bioinformatics* **2009**, *10*, 182.
[58] S. Perot, O. Sperandio, M. A. Miteva, A. C. Camproux, B. O. Villoutreix, *Drug Discov. Today* **2010**, *15*, 656-67.
[59] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, *Genome Res.* **2003**, *13*, 2498-504.
[60] Z. Zhang, M. G. Grigorov, *Proteins* **2006**, *62*, 470-8.
[61] G. Cruciani, M. Pastor, W. Guba, *Eur. J. Pharm. Sci.* **2000**, *11 Suppl 2*, S29-39.
[62] W. Xu, A. J. Lucke, D. P. Fairlie, *J. Mol. Graph. Model.* **2015**, *57*, 76-88.
[63] L. A. Pinna, F. Meggio, *Prog. Cell Cycle Res.* **1997**, *3*, 77-97.
[64] S. Sarno, S. Moro, F. Meggio, G. Zagotto, D. Dal Ben, P. Ghisellini, R. Battistutta, G. Zanotti, L. A. Pinna, *Pharmacol. Ther.* **2002**, *93*, 159-68.
[65] D. S. Wishart, C. Knox, A. C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, J. Woolsey, *Nucleic Acids Res.* **2006**, *34*, D668-72.
[66] A. P. Bento, A. Gaulton, A. Hersey, L. J. Bellis, J. Chambers, M. Davies, F. A. Kruger, Y. Light, L. Mak, S. McGlinchey, M. Nowotka, G. Papadatos, R. Santos, J. P. Overington, *Nucleic Acids Res.* **2014**, *42*, D1083-90.
[67] L. Dellafiora, M. Marchetti, F. Spyrakis, V. Orlandi, B. Campanini, G. Cruciani, P. Cozzini, A. Mozzarelli, *Bioorg. Med. Chem. Lett.* **2015**, *25*, 4297-303.
[68] G. Muratore, L. Goracci, B. Mercorelli, A. Foeglein, P. Digard, G. Cruciani, G. Palu, A. Loregian, .*Proc Natl. Acad. Sci. U. S. A.* **2012**, *109*, 6247-52.
[69] F. Spyrakis, P. Benedetti, S. Decherchi, W. Rocchia, A. Cavalli, S. Alcaro, F. Ortuso, M. Baroni, G. Cruciani, *J. Chem. Inf. Model* **2015**, *55*, 2256-2274.
[70] F. Spyrakis, B. Cellini, S. Bruno, P. Benedetti, E. Carosati, G. Cruciani, F. Micheli, A. Felici, P. Cozzini, G. E. Kellogg, C. B. Voltattorni, A. Mozzarelli, *ChemMedChem* **2014**, *9*, 1501-11.

[71] F. Spyrakis, R. Singh, P. Cozzini, B. Campanini, E. Salsi, P. Felici, S. Raboni, P. Benedetti, G. Cruciani, G. E. Kellogg, P. F. Cook, A. Mozzarelli, *PLoS One* **2013**, *8*, e77558.
[72] L. Goracci, S. Buratta, L. Urbanelli, G. Ferrara, R. Di Guida, C. Emiliani, S. Cross, *Eur J Med Chem* **2015**, *92*, 49-63.
[73] C. G. Fortuna, C. Bonaccorso, A. Bulbarelli, G. Caltabiano, L. Rizzi, L. Goracci, G. Musumarra, A. Pace, A. Palumbo Piccionello, A. Guarcello, P. Pierro, C. E. Cocuzza, R. Musumeci, *Eur J Med Chem* **2013**, *65*, 533-45.
[74] G. Muratore, B. Mercorelli, L. Goracci, G. Cruciani, P. Digard, G. Palu, A. Loregian, *Antimicrob Agents Chemother* **2012**, *56*, 6009-13.
[75] L. Goracci, M. Ceccarelli, D. Bonelli, G. Cruciani, *J Chem Inf Model* **2013**, *53*, 1436-46.
[76] D. Cardinale, G. Guaitoli, D. Tondi, R. Luciani, S. Henrich, O. M. Salo-Ahen, S. Ferrari, G. Marverti, D. Guerrieri, A. Ligabue, C. Frassineti, C. Pozzi, S. Mangani, D. Fessas, R. Guerrini, G. Ponterini, R. C. Wade, M. P. Costi, *Proc Natl Acad Sci U S A* **2011**, *108*, E542-9.
[77] E. Carosati, G. Sforna, M. Pippi, G. Marverti, A. Ligabue, D. Guerrieri, S. Piras, G. Guaitoli, R. Luciani, M. P. Costi, G. Cruciani, *Bioorg Med Chem* **2010**, *18*, 7773-85.
[78] R. Z. Cer, U. Mudunuri, R. Stephens, F. J. Lebeda, *Nucleic Acids Res.* **2009**, *37*, W441-5.

“For Table of Contents use only”

Comparing drugs image and repurposing drugs with BioGPS and FLAPdock: the Thymidylate Synthase Case

*Lydia Siragusa [1], Rosaria Luciani [2], Chiara Borsari [2], Stefania Ferrari [2], Maria Paola Costi [2], Gabriele Cruciani [3], Francesca Spyrakis [2,4,]**



We describe the integrated and innovative BioGPS/FLAPdock pipeline for the rapid and effective comparison of protein cavities, off-targets identification and drug repurposing. Structural, chemical and energetic properties of the pockets are simply encoded in the corresponding GRID Molecular Interaction Fields. BioGPS discloses pockets similarity and cross-docking experiments identify drugs to be possibly repurposed. The Thymidylate synthase test case is reported.