

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

Lydia Siragusa,^[b] Rosaria Luciani,^[a] Chiara Borsari,^[a] Stefania Ferrari,^[a] Maria Paola Costi,^[a] Gabriele Cruciani,^[c] and Francesca Spyrakis*^[a, d]

Repurposing and repositioning drugs has become a frequently pursued and successful strategy in the current era, as new chemical entities are increasingly difficult to find and get approved. Herein we report an integrated BioGPS/FLAPdock pipeline for rapid and effective off-target identification and drug repurposing. Our method is based on the structural and chemical properties of protein binding sites, that is, the ligand image, encoded in the GRID molecular interaction fields (MIFs). Protein similarity is disclosed through the BioGPS algorithm by measuring the pockets' overlap according to which pockets are clustered. Co-crystallized and known ligands can be cross-

docked among similar targets, selected for subsequent in vitro binding experiments, and possibly improved for inhibitory potency. We used human thymidylate synthase (TS) as a test case and searched the entire RCSB Protein Data Bank (PDB) for similar target pockets. We chose casein kinase II α as a control and tested a series of its inhibitors against the TS template. Ellagic acid and apigenin were identified as TS inhibitors, and various flavonoids were selected and synthesized in a second-round selection. The compounds were demonstrated to be active in the low-micromolar range.

Introduction

With the threat of a 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research has to deal with resistance to chemotherapy and molecular targeted therapies^[4] and, in general, drugs are withdrawn from the market or during clinical phases because of toxic effects.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 far from Ehrlich's magic bullet,^[5] the development and discovery of drugs that are able to bind more than one single target, acting as a magic shotgun,^[6] is very desirable. Polypharmacology has emerged as the next paradigm of drug discovery,^[7–12] and off-target prediction is a key issue in bioinformatics and drug design. The identification of unknown targets can

not only lead to unexpected applications or to multi-target drugs, but also to prevent adverse effects from occurring in clinical trials, or worse, when drugs have already reached the market. In addition, drug repurposing (using already approved and safe drugs for new targets), would save money, time and resources. A number of methodologies have appeared in recent years with the aim of identifying off-targets, predicting side effects (SEs) and possibly finding new applications for already known molecules.

Notable efforts have been made to develop in vitro assays and standardize procedures 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 with in vitro analyses. Different methodologies have been developed for off-target identification and drug repositioning.^[14,15] While initial 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 correlation analyses were thus applied to link the ligand chemical space to the targets and to the possible related 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

[a] Dr. R. Luciani, Dr. C. Borsari, Dr. S. Ferrari, Prof. M. P. Costi, Dr. F. Spyrakis
Department of Life Sciences, University of Modena and Reggio Emilia,
Via Campi 103, 41125 Modena (Italy)
E-mail: francesca.spyrakis@unipr.it

[b] Dr. L. Siragusa
Molecular Discovery Limited,
215 Marsh Road, Pinner Middlesex London HA5 5NE (UK)

[c] Prof. G. Cruciani
Department of Chemistry, Biology and Biotechnology,
University of Perugia, Via Elce di Sotto 8, 06123 Perugia (Italy)

[d] Dr. F. Spyrakis
Current address: Department of Food Science, University of Parma, Viale
delle Scienze 17A, 43124 Parma (Italy)

Supporting information [a list of the 317 selected cavities along with their PDB and UniProt codes and co-crystallized ligands; a detailed description of the pocket connections and Volsurf-based analysis of the pockets; and docking poses of active flavonoids within TSpw] and the ORCID identification number(s) for the author(s) of this article can be found under <http://dx.doi.org/10.1002/cmdc.201600121>.

and target information have been reported showing that a drug action is often unspecific, and underlying the necessity of combining biology and chemistry to provide reliable molecular explanations for complex SEs.^[34] Ligand binding site comparison and protein–ligand docking have been also successfully integrated and applied for drug repurposing, side effect prediction and polypharmacology applications.^[35–38]

Approaches directly comparing protein pockets have also been recently proposed, based on the assumption that similar binding sites can be targeted by similar ligands,^[39–43] and that the structural and chemical information encoded into the binding pockets guide the recognition between macromolecules and ligands.^[44–47] SMAP is a fast method for ligand binding site comparison, using a shape description only based on C α atoms.^[48] SiteEngine^[45] Cavbase,^[44] and FuzCav^[46] represent each pocket residue as a series of pseudocenters, encoding the physicochemical properties essential for molecular interactions. These rule-based methods generally produce models in which all pseudocenters are equally considered, without paying attention to the residue environment. Slightly different strategies are applied by PocketFEATURE, which compares protein sites by one or more microenvironments, described as occurrence of atoms, residues, and biochemical and biophysical properties,^[49] and by ProBIS, which is able to detect structurally similar sites as patterns of physicochemical properties on the protein surface.^[50]

In this mazy plethora of options, structure-based methods can provide mechanistic indications about the off-targets selection and about the related occurrence of side effects. Moreover, the possibility of finely tuning the effect of an approved or candidate drug toward 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 interactions, we developed the BioGPS (Global Positioning System in Biological Space) algorithm,^[51] based on the chemical/structural comparison of protein binding sites.^[51,52] BioGPS represents and compares pockets according to their ligand image and not by any other rule-based residue feature.^[44] Pockets are described by their molecular interaction fields (MIFs, calculated by GRID^[53,54]), that is, the shape, the hydrophobic regions, the H-bond donor and acceptor hotspots 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.^[55] No sequence-related information, ligand similarity or side effect 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 with a possible interaction with these groups. The larger the MIF produced by one or more residues, the higher the probability to find a complementary group in that region, i.e., a 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 respectively. Comparing the MIFs means comparing the chemical and geometrical properties as well as the encoded energetics. Recent applications have demonstrated BioGPS' capability of predicting off-target effects, classifying protein families, justifying polypharmacology, and rationalizing selectivity between sub-families.^[51,52,55]

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 RCSB Protein Data Bank (PDB), 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.^[56]

The pipeline is composed of four main steps: 1) data collection, 2) cavities comparison and selection, 3) ligand docking, and 4) in vitro binding experiments.

1) *Data collection.* The first step consists of the selection of the protein template 3D structures and of the database of proteins to compare with it. For each protein structure the co-crystallized ligand and the cavity containing it (the binding site) are detected.

2) *Cavities comparison.* The template pocket is compared with all of the other pockets by using the BioGPS algorithm. Cavities in the database are ranked according to their similarity

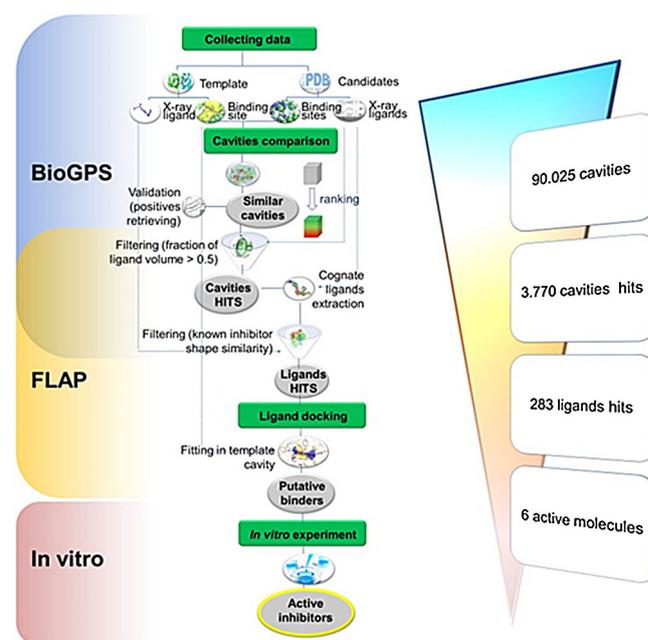


Figure 1. Workflow of the drug-repurposing approach based on the BioGPS/FLAPdock integrated technology.

to the template. The algorithm's capability of retrieving pockets belonging to the same template protein family is evaluated by means of enrichment analysis. An established similarity threshold (Global Product >0.7) is used for selecting similar cavities. To avoid sites not completely occupied by ligands, only cavities containing at least 50% of the co-crystallized ligand volume are retained. Once cavity hits are selected, co-crystallized ligands are extracted and filtered according to their volume similarity with respect to known template inhibitors.

3) *Ligand docking*. The extracted ligands are docked within the template binding site with the FLAPdock algorithm. The 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.

4) *In vitro binding experiments*. We used human thymidylate synthase (TS) as a template. Human and bacterial TS catalyzes 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 methyl group to dUMP, thus transforming it into dTMP.^[57] 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 follows that human TS represents a good pharmacological target for anticancer drugs and antimicrobials. Nucleotide- and folate-like inhibitors are, in fact, used in cancer chemotherapy because of the cytotoxic effects of thymidylate depletion.^[58-60] Microbial TSs have also been demonstrated to be suitable targets for antimicrobial agents.^[61,62]

The TS binding site was used as a template to screen the entire PDB, looking for similar cavities. Human and bacterial TS were reasonably identified as the most similar proteins, being followed by apparently diverse candidates such as kinases, proteases, phosphodiesterases, nuclear receptors, and chaperones, among 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 illustrate the strength of this approach in that it is able to automatically and quickly identify similarity between the same and different protein families.

To identify 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 the TS cavity. Interestingly, the most promising molecules were inhibitors of casein kinase II α (CKII α), which emerged as a TS off-target. Two co-crystallized ligands and a series of related flavonoids were tested in vitro for inhibition activity toward TS. Six compounds inhibited TS in the low micromolar range, thus supporting the potential of the BioGPS approach for drug repurposing campaigns.

Results and Discussion

Thymidylate synthase can be present in various configurational states: 1) the apo inactive form, 2) an inactive form complexed with peptides binding the homodimer interface, 3) an active form complexed with dUMP, and 4) an active form complexed with dUMP and mTHF or an antifolate drug (ternary complex). TS undergoes significant conformational rearrangements upon dUMP binding. In particular, in the native unbound form the 181–197 loop, containing the catalytic cysteine, is rotated $\sim 180^\circ$ 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.^[63] Looking for new unknown TS ligands, we focused on the protein closed active conformation in presence of the dUMP substrate (Figure 2).

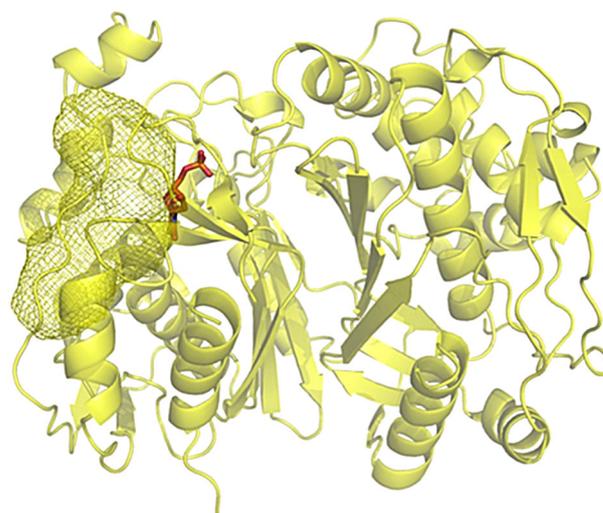


Figure 2. Ribbon representation of human TS. The binding site of one monomer is represented as yellow mesh lines; dUMP is shown as orange capped sticks.

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, that is, PDB ID: 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 the TS binding site.^[63] 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 interac-

tion 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 GRID OH2 probe energy of the aforementioned waters within the binding pocket. The two selected molecules showed an energy value lower than -8 kcal mol^{-1} , a strong evidence of their importance in complex formation and stabilization. Thus, two different human TS structures were finally modelled: 1) TS with dUMP and without any water molecules (TSp) and 2) 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 templates to perform the following screening and docking experiments, with the perspective that the incoming ligands might displace any water (TSp) or retain and exploit the existing ones (TSpw).

First step: screening the pockets

The two TS pockets were used as templates to screen the PDB with BioGPS. In particular, 90025 pockets were screened, which correspond to all 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 GRID MIFs. Template MIFs were compared with cavity database MIFs, calculating for each pairwise comparison a set of nineteen FLAP scores, representing the similarity of the match (see the Experimental Section for further details on BioGPS and FLAP). In particular, the Global Product (GlobP), the product of the four principal FLAP scores, was used to evaluate the degree of similarity between the templates and the candidate cavities. The GlobP score ranges from 0, for no superposition, to 1, for a perfect pocket overlap complete with identical interactions, and provides a global evaluation of both geometric and chemical similarity. Thus, the 90025 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.

The plot suggests that the water molecules enhance the promiscuity of the cavity. Indeed, using TSpw as a 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 a 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 90025 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.^[51]

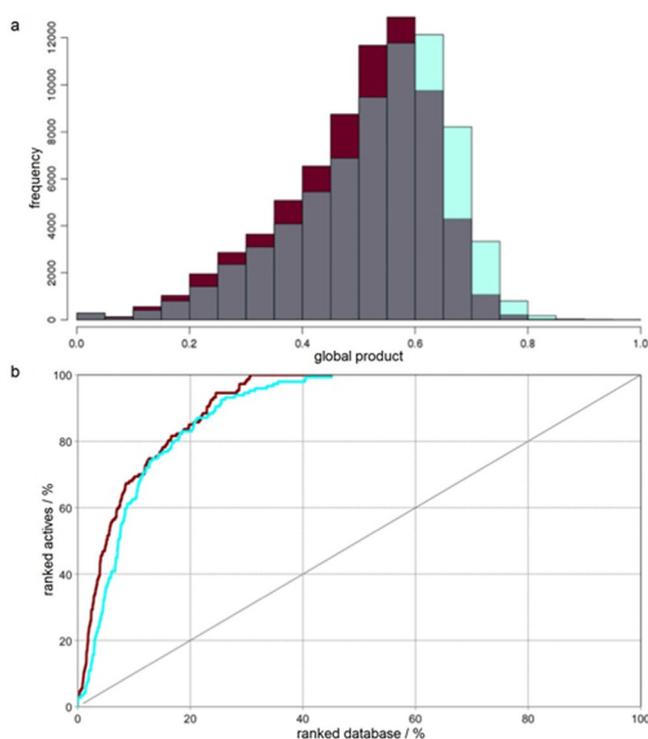


Figure 3. BioGPS virtual screening statistics. a) Distribution of the similarity Global Product score calculated by the BioGPS virtual screening. TSp and TSpw, colored red and cyan respectively, were used as template pockets. Overlap areas are colored grey. b) Enrichment curves calculated for the two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

1336 cavities, relating to 606 unique proteins, and 4349 cavities, relating to 1513 unique proteins, were selected as hits for TSp and TSpw, respectively (Figure 4). 1208 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 of the 4476 identified pockets, corresponding to 1536 proteins. To further

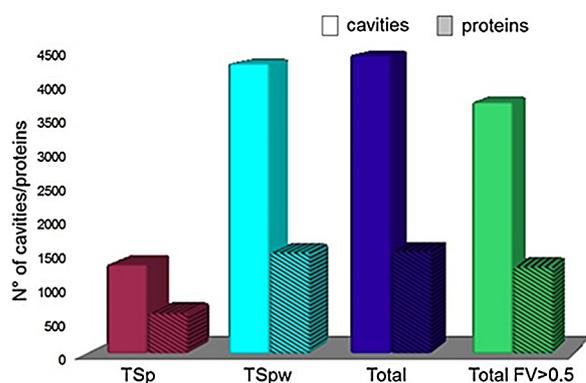


Figure 4. Cavity hits, and corresponding proteins, identified by the BioGPS VS. The hits identified for the TSp and the TSpw templates are respectively colored purple and cyan. The total number of hits is colored blue, while the green bars correspond to the hits number filtered for the ligand fraction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

decrease this large number 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.^[64,65] 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, with some of them present in different pockets. Irrelevant compounds such as solvents, that is, ethylene glycol, glycerol, or prosthetic groups were discarded. The remaining ligands were compared, in terms of volume, to known human TS inhibitors retrieved from the ChEMBL database (see The Experimental Section for further details). Those having a volume higher than the smallest TS ligand and lower than the largest one were retained, to consider compounds in the same volume range of known TS inhibitors and in the same druggable space. This led to a decreased set of 283 ligands belonging to 135 proteins and 317 pockets (Table S1).

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 analysis on the 317 hits, constructing a pairwise similarity matrix. We then built a binding site similarity network using Cytoscape,^[66,67] where each node represents a pocket and pockets are connected and clustered if they have a GlobP higher than 0.4. The higher the GlobP for a cavity pair, the closer the two cavities will be in the graph, that is, more similar from a structural, chemical and energetic point of view.

The high color/cluster correspondence reported in Figure 5 shows how BioGPS is not only able to detect pocket similarities, but also to cluster proteins according to their binding site properties. In our study pockets are only described in terms of four molecular interaction fields, that is, the shape, the hydrophobicity, the H-bond donor and acceptor character. These seemingly simple properties that encode the energetic description of the cavities through GRID force field and the subsequent FLAP GlobP global similarity description 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 classify proteins into their corresponding families. 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 chaperones or nuclear receptors and oxidoreductases, but with more difficulty among chaperones and oxidoreductases. Chaperone ligands could also be

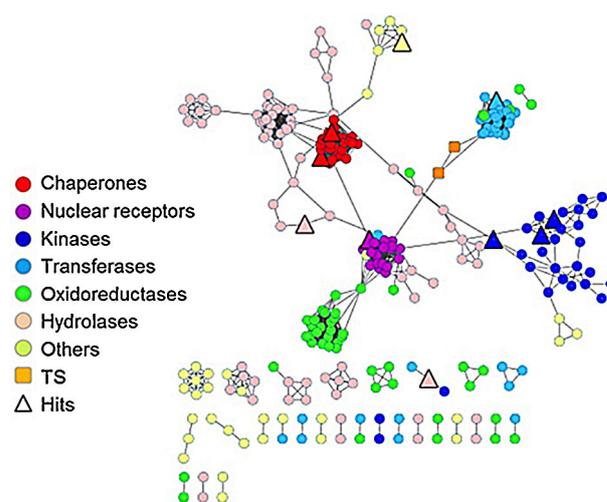


Figure 5. BioGPS-network. Similarity network of the 317 selected cavities built according to the cavities pair Global Product estimated by BioGPS. Each node represents a cavity and each edge between two nodes represents the similarity between them (the shorter the distance, the higher the similarity). The different protein classes are color-coded according to the legend. Singletons are reported below. Hit proteins, that is, containing the ten selected ligands reported in Table 1, Figure 6, are identified by triangles.

potentially repurposed for phosphodiesterases (PDEs) and *vice versa*, while it would be harder, in principle, to relocate PDE ligands in nuclear receptors (see the Supporting Information for a detailed description of the pocket distribution).

From 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. We further characterized and separated the cavities according to their morphological and chemical properties using Volsurf.^[68] Results are reported in the Supporting Information (Figure S1).

Second step: docking cognate ligands

From the 317 cavities we retrieved 283 co-crystallized different ligands (Table S1). The ligand dataset was prepared considering tautomers and protomers using MoKa and docked within TSp and TSpw with FLAPdock, the docking tool implemented in FLAP.^[56] About 80% of the compounds was first removed according to the FLAP S-score. Only molecules with a score value higher than 0.90 were retained, based on previous observations [data not shown]. The retained compounds were visually inspected and further filtered according to the complementarity of the pocket MIFs with the ligand pseudo-MIFs, and to the number of hydrogen bonds formed with the residues lining the cavity. According to the literature and to studies we previously performed, the visual inspection of the results by trained operators can, indeed, strongly improve the success rate of screening and docking experiments.^[69–72]

The ten most promising molecules were selected and are reported in Figure 6 and Table 1. Ligands were retrieved from casein kinase II α , LFA-1 integrin, heat shock protein 90, β -secre-

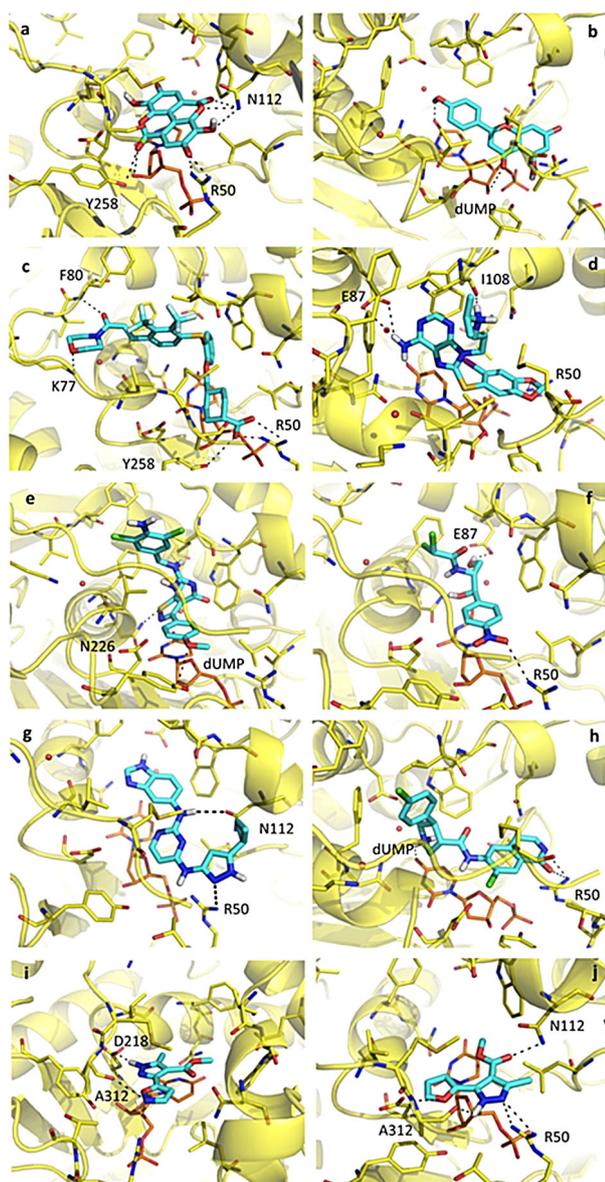


Figure 6. Docking poses of the ten best selected compounds within the hTS binding pockets as predicted by FLAPdock. The PDB ligand code and the ligand name are reported as follows: a) REF, ellagic acid, 2,3,7,8-tetrahydroxy-chromeno[5,4,3-*cde*]chromene-5,10-dione; b) AGI, apigenin, 5,7-dihydroxy-2-(4-hydroxyphenyl)-4*H*-chromen-4-one; c) E2M, *cis*-4-[[2-[(4-[(1*E*)-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]-9*H*-purin-6-amine; e) OVA, *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*-[(1*R*,2*R*)-2-hydroxy-1-(hydroxymethyl)-2-(4-nitrophenyl)ethyl]acetamide; g) APJ, *N*²-1*H*-benzimidazol-5-yl-*N*⁴-(3-cyclopropyl-1*H*-pyrazol-5-yl)pyrimidine-2,4-diamine; h) 3ND, (3*S*,4*R*)-*N*-(7-chloro-1-oxo-1,4-dihydroisoquinolin-6-yl)-4-(4-chlorophenyl)pyrrolidine-3-carboxamide; i) 2D3, methyl 3-isoxazol-5-yl-5-methyl-1*H*-pyrazole-4-carboxylate; j) 37D, methyl 5-furan-2-yl-3-methyl-1*H*-pyrazole-4-carboxylate. dUMP is shown in orange color sticks. Residues involved in hydrogen bonding the ligands are labeled. When present, water molecules are displayed as red spheres. Complexes a)–h) were obtained by docking in TSpw, complexes i), j) by docking in TSp.

tase 1, chloramphenicol acetyltransferase 3, serine/threonine-protein kinase/endoribonuclease IRE1 and Rho associated protein kinase.

The ligands, quite different from each other in terms of shape, volume and chemical properties, present, predictably, diverse interactions with the pocket residues. Only the π - π contact with dUMP is maintained overall, with the exception of E2M ligand (Figure 6c). 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 hydrogen bonded to a pyrimidine carbonyl and to the hydroxy group of the deoxy-ribose dUMP. In addition, slightly different apigenin poses within the TS pockets show 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 π - π interaction with the dUMP pyrimidine moiety (Figure 6a,b). The other molecules also mainly contact Arg50, Asn112, Tyr258 and dUMP. Additional hydrogen bonds are formed with Phe80, Glu87, Ile108, Asp218, Asn226, Ala312 (Figure 6).

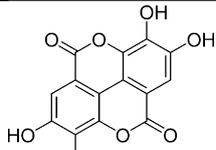
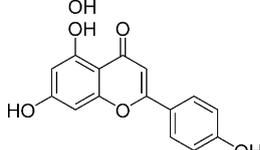
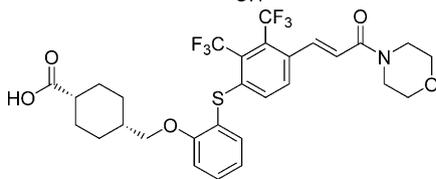
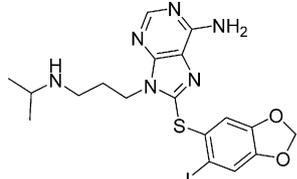
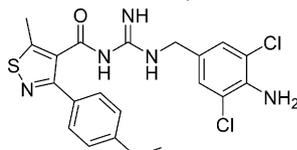
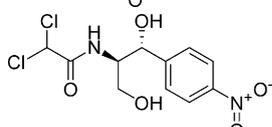
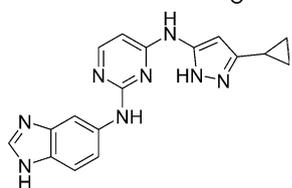
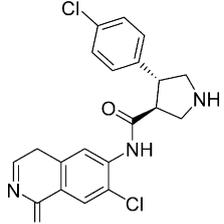
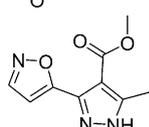
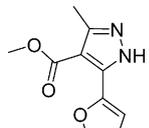
Interestingly, both ellagic acid and apigenin are known inhibitors of CKII α . CKII is an attractive anti-neoplastic and antiviral target, essential for cell viability and with many cellular targets.^[73] The catalytic subunits of CKII, α and α' , are constitutively active, either alone or in combination with the β subunit, a necessary property for the continuous need to phosphorylate its numerous targets, but also potentially dangerous in neoplastic pathologies and viral infections.^[74]

Biological evaluation of ligands and synthesis of second-round compounds

As mentioned, ellagic acid and apigenin were the most interesting compounds, and are both inhibitors of CKII α , having $IC_{50} = 40$ nM and $K_i = 740$ nM, respectively.^[75] We therefore experimentally evaluated them for inhibition toward 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 a K_i value of 16 μ M, while apigenin only showed poor inhibition of 6% at 100 μ M. Considering that apigenin did not fulfill the H-bond potential of TS pockets (Figure 6b) and that flavonoids are known CKII inhibitors, we extended the analysis to other purchased and in-house synthesized flavonoids (compounds 1–4, Table 2, second part; see The Experimental Section 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 of compound selection, similar to those 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.

Table 1. Structures and properties of the most promising compounds selected by the BioGPS/FLAPdock integrated approach.

PDB code	Structure	Co-crystallized protein (PDB ID)
REF		casein kinase II α (2ZJW)
AGI		casein kinase II α (3AMY)
E2M		LFA-1 binding domain (3E2M)
H71		Hsp90 (2FWZ)
OVA		β -secretase 1 (4FSE)
CLM		chloramphenicol acetyltransferase 3 (4CLA)
APJ		Ser/Thr protein kinase/endoribonuclease IRE1 (3FBV)
3ND		Rho-associated protein kinase (3NDM)
2D3		Hsp90 (2YE8)
37D		Hsp90 (3HZ1)

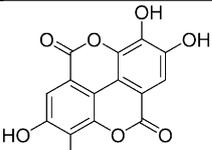
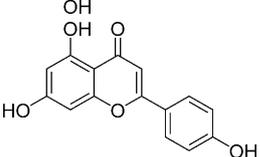
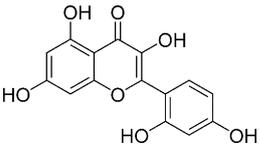
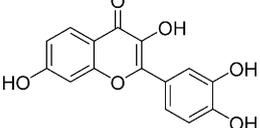
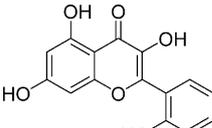
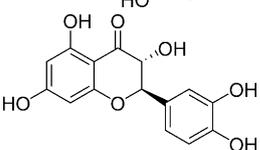
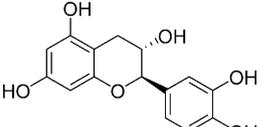
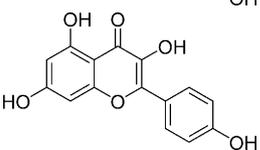
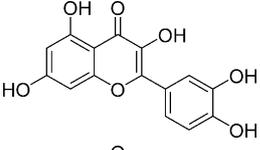
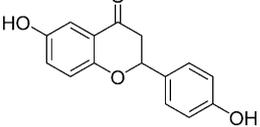
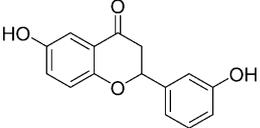
Name	M_r [Da]	Structure	IC_{50}/K_i [μM] ^[a,b]
ellagic acid	302.19		238/16
apigenin	270.24		1440/94
morin	302.24		45/2.9
fisetin	286.24		52/3.4
datisctein	286.24		NI
taxifolin	304.25		NI
catechin	290.27		NI
kaempferol	286.24		57/3.7
quercetin	302.24		NI
compound 1	256.25		101/6.6
compound 2	256.25		NI

Table 2. (Continued)			
Name	M_r [Da]	Structure	IC_{50}/K_i [μM] ^[a,b]
compound 3	272.25		273/18
compound 4	286.24		61/4.0

[a] Standard error within $\pm 20\%$ of the given value.^[61] [b] NI: no inhibition at the solubility limit.

In particular, seven were purchased from vendors and four (compounds 1–4) were synthesized in house, as reported in Scheme 1. Among the acquired compounds, morin, fisetin, and kaempferol presented an inhibition effect in the low micromolar range (K_i : 2.9, 3.4, 3.7 μM , respectively). Compounds 1 and 4 showed a $K_i < 10 \mu M$ (K_i of 6.6 and 3.3 μM , respectively), compound 3 a poor inhibition of 8% at 25 μM , while 2 could not be studied because of the low solubility. According to ChEMBL, known TS inhibitors present K_i values ranging from low nanomolar to high micromolar values.^[75–77] Our hits, having, in general inhibition values in the low micromolar range represent good potential starting point for the development and optimization of more potent inhibitors. 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 10–100 μM range.

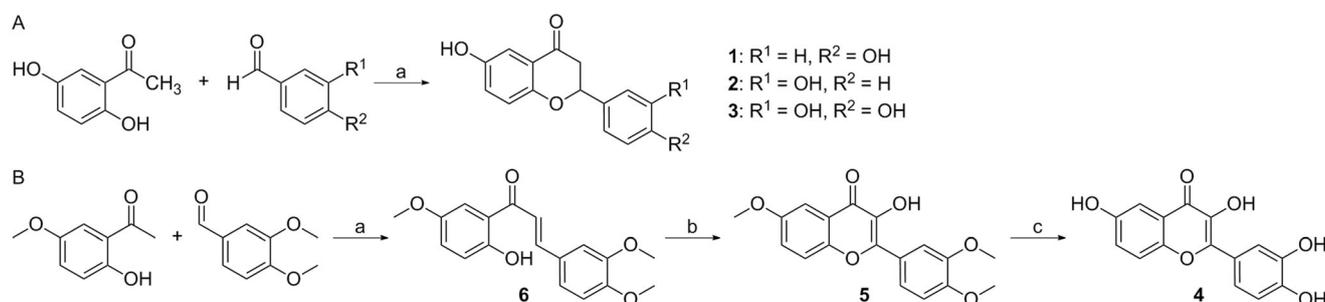
Docking poses of the active compounds within the TS binding site are reported in Figure S2. The presence of more polar groups on the molecules allowed the formation of additional hydrogen bonds within the target pocket, which might explain their higher activity. While apigenin was able to form a few hydrogen bonds (in the pose shown in Figure 6b only the dUMP substrate is contacted) the five active flavonoids interact with Arg 50, Glu 87, Asn 112, Asp 218, Val 223, Tyr 258 and Ala 312. This higher protein–ligand structural complementarity 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 compound selection and optimization.

Both ellagic acid and apigenin are known inhibitors of other proteins apart from CKII. According to Drugbank,^[78] ellagic acid is active against carbonic anhydrase 1, 2, 3, 4, 5A mitochondrial, 5B mitochondrial, 6, 7, 9, 12, 14, CKII α , cAMP-dependent protein kinase α , protein kinase α and β 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.^[75] Among these we find aldehyde dehydrogenase 1A1, cytochrome P450 2C9, 2C19, 2D6 and 3A4, DNA polymerases, ERBB1, MAP kinase ERK2 and p38 α , acetylcholinesterase, cholinesterase, β -secretase 1, pyruvate kinase, protein kinase C α , Rho-associated protein kinase 2, tyrosine-protein kinase SYK, phosphodiesterase 5A, and others. Despite their rather promiscuous profile, for both ellagic acid and apigenin TS has not been previously reported as a potential target. In addition morin, fisetin and 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 CKII α pockets present similar properties. We compared the pockets MIFs and reported their superimposition in Figure 7. The pockets shape and dimension present a certain degree of similarity, with vol-



Scheme 1. A) Synthesis of compounds 1–3: a) $SOCl_2$, EtOH, RT; B) Synthesis of compound 4: a) NaOH (3 M), EtOH, RT; b) H_2O_2 , NaOH (1 M), EtOH, RT; c) BBr_3 (1 M in dry DMC), dry DMC, $0^\circ C \rightarrow RT$.

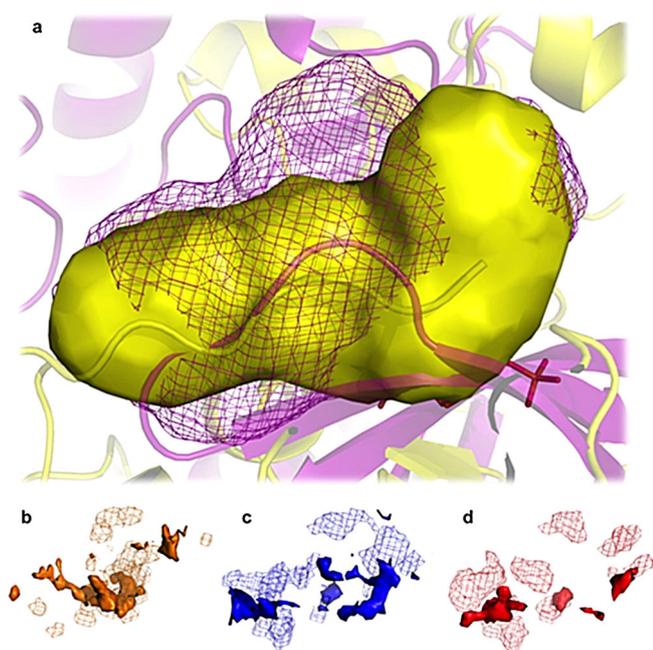


Figure 7. Superimposition of hTS and CKII binding sites. a) Shape superimposition, TS pocket (PDB ID: 1HVY) is shown as yellow surface and CKII α (PDB ID: 2ZJW) pocket as magenta mesh lines. b),c),d) Superimposition of hydrophobic, H-bond donor and H-bond acceptor molecular interaction fields, respectively; TS MIFs are shown as solid surfaces while CKII α MIFs as mesh lines.

umes of 1390 and 1675 Å³, for TS and CKII α , respectively (Figure 7a). 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 π - π interactions with the dUMP in TS. A lower superimposition can be observed for the H-bond donor and H-bond acceptor group. Clearly, a total overlap of pockets' chemical and geometrical properties is not expected, 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 ER α and the SERCA (sarcolemmal reticulum Ca²⁺ ion channel ATPase) cavity, which is a known off-target for selective estrogen receptor modulators.^[79] The two superposed cavities showed only a 62% overlap of the volume. Again the highest similarity scores were detected for the hydrophobic and the H-bond donor MIFs.^[51] Also it must be remembered that proteins and ligands are flexible and so is the image that ligands produce in their binding site. There is no expectation that ligands will be totally complementary toward the new target, but recognize at least a part of the image they produced in the original pocket.

Conclusions

Recognizing the potential of drug repurposing strategies, we have presented here the BioGPS/FLAPdock approach for off-target identification and repositioning applications. 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 entering into a specific pocket. Pockets are represented and compared according to their GRID molecular interaction fields, also encoding the energetics of the pockets, differently from many other approaches. Once similar pockets are identified, co-crystallized ligands, or any other known inhibitor can be cross-docked between the template and the queries, or among the related queries to look for new targets and applications.

We ran this pipeline for the specific TS case, identifying Hsp90, the estrogen receptor, the vitamin D3 receptor, different kinases, transferases and phosphodiesterases and others as possible related targets. We selected some ligands from CKII α , ellagic acid and flavonoids, and found them to be inhibitors of the TS template in the low micromolar range.

The specific case described here 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 illustrate 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. After the selection of possible related targets, ligands can be exchanged by docking simulations, then tested and improved by means of *in silico* analysis, *in vitro* analysis, and chemical synthesis. This second step allows the rationalization of the protein–ligand interaction, the potential 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 uses. Pocketomes can be easily and rapidly analyzed for identifying targets likely responsible for unpredicted side effects. Further, 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's 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 from the pocket definition (which is itself automatically defined). The simplicity of the pocket search, the rapid and semi-automated procedure, makes it a promising and valuable tool for modeling polypharmacology, drug repurposing and side effects.

Experimental Section

Cavity identification: The FLAPsite algorithm is used for the identification of cavities in 3D protein structures.^[51] By embedding the protein structure into a 3D grid with a spatial resolution of 1.0 Å, the algorithm identifies pocket points using the GRID probe H (shape).^[53] 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 ap-

plied to the remaining points for removing small anomalies and connecting areas. The hydrophobic probe DRY is used to prioritize hydrophobic cavities usually targeted by drugs. The FLAPsite procedure was applied to 1) human TS (PDB ID: 1HVY) with dUMP, without any water molecule (TSp); 2) human TS (PDB ID: 1HVY) with dUMP and two bridging water molecules (wat435 and wat622; TSpw); 3) all PDB protein structures co-crystallized with a ligand. Only cavities containing a ligand were selected for the following steps of the pipeline: (1–2) two TS template cavities and (3) 90025 dataset cavities.

Virtual screening with BioGPS: We collected from the PDB the structures of all proteins co-crystallized 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 then detected by using the FLAPsite algorithm (90025 binding sites; September 2014). MIFs were calculated for each binding site and stored in a database. The BioGPS technology was used to compare two TS pocket templates against 90025 MIFs cavities dataset. The BioGPS algorithm compares binding sites by means of their MIF similarity,^[51,55] and exploits the technology implemented in FLAP.^[56] 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,^[69,70,80–85] as well as binding mode prediction and rationalization.^[86,87] 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. Because a simple comparison of the entire MIF areas might be computationally expensive, the algorithm decreases the information by selecting a number of representative points, called hotspots, proportional to the energy and the volume of each MIF. All possible combinations of four hotspots (called quadruplets) are generated and stored in a fingerprint, named Common Reference Framework.^[56] The BioGPS algorithm compares two cavities by comparing such Common Reference Frameworks in a pairwise manner. This approach searches for the largest number of favorable quadruplet superpositions. 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 $<1 \text{ \AA}$), the algorithm overlaps their 3D cavity structures with a specific orientation, according to the matching quadruplet, and then calculates the MIF similarity of the overlapping areas. The final superposition is called a “solution” and is quantitatively scored by considering the corresponding MIFs similarity, summarized 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, that is, 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, in-

cluding the Global Sum and Global Product, FLAP also calculates a Distance Score, representing the overall similarity derived from a combination of all calculated similarity scores computed for the candidates and the template, that is, the protein binding site. The Global Product score was considered for hit selection in this study; it ranges between 0 and 1, where the higher the score the more similar are the two entities.

Hit (pockets) selection: The Global Product score was set to 0.7 as restrictive threshold for selecting the most similar cavities to the hTS templates; 4476 cavities were selected. Then, for each cavity the FV was calculated as the ratio of the ligand volume contained within a cavity over the total volume of the ligand. Only cavities with $\text{FV} > 0.5$ were selected. We thus obtained 3770 cavities, from which 1361 co-crystallized ligands were extracted. Irrelevant compounds such as solvents, that is, ethylene glycol, glycerol, or prosthetic groups were discarded. The remaining ligands were compared with a set of known human TS inhibitors, retrieved from the ChEMBL database.^[75] A bioactivity search was performed and 354 inhibitors with known K_i toward 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. Their minimum and maximum volumes were used as cut-offs. Unknown ligands (extracted from cavities) having a volume higher than the smallest hTS inhibitor and lower than the biggest one were retained. This process yielded 283 ligands belonging to 135 proteins and 317 pockets (Table S1). We thus confined the research to ligands occupying the same druggable space of known TS inhibitors. This filter was applied to increase the probability of success and to create a focused and reduced dataset to perform in vitro analyses and validate the pipeline. Discarded ligands will be further analyzed to investigate the effect of enlarging the chemical space on the compounds activity.

Connectivity analysis: Network graphs reported in Figure 5 and S1 were built with Cytoscape.^[66] 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 against the others. Because 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 $1 - \text{GlobP} < 0.6$, that is $\text{GlobP} > 0.4$. 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 $\text{GlobP} > 0.7$ with respect to TS), stricter thresholds were set. $1 - \text{GlobP} < 0.2$ ($\text{GlobP} > 0.8$) and $1 - \text{GlobP} < 0.05$ ($\text{GlobP} > 0.95$) were the thresholds considered to connect TS with the other pockets and to itself, respectively.

Molecular docking: FLAPdock is a docking approach implemented in the software FLAP,^[56] based on GRID MIF similarities,^[53] 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 minimal 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 ener-

getic terms for this subset; the solutions are then ranked according to the combined score, 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 ultimately 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 pseudo-MIFs. Pseudo-MIFs correspond to the projection of the MIFs on the atoms that generate them. Ligand tautomer and protomer enumeration was performed with MoKa 2.6 prior to the docking simulations. Docking analyses were performed within both TSp and TSpw.

Protein expression and purification: Human TS was cloned in the pQE80 L system, as reported.^[88] 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 through immobilized metal affinity chromatography. Bacteria (DH5a/pQE80L) were grown in LB medium containing (2 L) ampicillin (50 mg mL⁻¹). The solution was centrifuged at 37 °C, 120 rpm, until the OD_{600nm} reached a value of 0.6. TS expression was induced adding isopropyl-β-D-thiogalactopyranoside (IPTG, 1 mM) and incubating the culture for 4 h, 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 NaH₂PO₄, 30 mM NaCl, 20 mM imidazole, pH 7.5) containing Complete® (protease inhibitor) and sonicated in an ice bath. The broken cells were centrifuged for 40 min at 12000 rpm, 4 °C, and the pellet was discarded. The supernatant was treated with streptomycin (10%), stirred for 10 min at 4 °C and centrifuged for 30 min, 12000 rpm, 4 °C. The discarded pellet and the supernatant were filtered (0.8/0.45 μm filters) and loaded on a Ni-NTA column pre-equilibrated with buffer A. The enzyme was eluted with buffer B (20 mM NaH₂PO₄, 30 mM NaCl, 1 M imidazole, pH 7.5). The fractions with enzyme were collected, pooled and loaded on a HiTrap Desalting column to change the buffer with buffer C (20 mM NaH₂PO₄, 30 mM NaCl, pH 7.5).^[89] 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 min 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 (0.04–0.3 μM). The reaction mixture contained 50% of assay buffer (TES, *N*-[tris(hydroxymethyl)methyl]-2-aminoethanesulfonic acid (100 mM), MgCl₂ (50 mM), formalin (13 mM), EDTA (2 mM), pH 7.4, β-mercaptoethanol (150 mM)), the enzyme (0.1 μM), mTHF (50 μM), dUMP (120 μM) 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 (10 mM) were solubilized in DMSO.^[89] The inhibition percentage was determined upon evaluation of the differential optical depth (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. Samples

(100, 50, 25 or 10 μM) were prepared, decreasing 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 of gradually increasing the compounds' concentration and experimentally determining the IC₅₀ value, IC₅₀ and *K_i* were calculated from the inhibition percentage as reported.^[90] Values might be underestimated.

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. ¹H and ¹³C 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; brs, 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 1. 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 equiv) in absolute EtOH (2 mL), thionyl chloride (120 μL) was added dropwise over 5 min. The reaction was stirred at room temperature for 6 h. EtOH and excess 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 with 31% yield; mp: 230 °C. ¹H NMR (CD₃OD, 400 MHz): δ = 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 ppm (dd, 1H, *J*_{3a,3b} = 17.0 Hz, *J*_{3a,2} = 2.8 Hz, Ha-3); ¹³C NMR (CD₃OD, 100 MHz): δ = 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 ppm; ESI-HRMS calcd for C₁₅H₁₃O₄ [M + H]⁺ 257.0808, found 257.0805.

6-Hydroxy-2-(3-hydroxyphenyl)chroman-4-one (2): was isolated as a yellow solid with 26% yield; mp: 240 °C. ¹H NMR (DMSO, 400 MHz): δ = 9.51 (brs, 1H, 3'-OH), 9.42 (brs, 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 ppm (dd, 1H, *J*_{3a,3b} = 16.8 Hz, *J*_{3a,2} = 2.6 Hz, Ha-3); ¹³C NMR (DMSO, 100 MHz): δ = 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 ppm; ESI-HRMS calcd for C₁₅H₁₃O₄ [M + H]⁺ 257.0808, found 257.0808.

2-(3,4-Dihydroxyphenyl)-6-hydroxychroman-4-one (3): was isolated as an orange solid with 26% yield; mp: 220 °C. ¹H NMR (DMSO, 400 MHz): δ = 9.35 (brs, 1H, OH), 8.90 (brs, 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 ppm (dd, 1H, $J_{3a,3b}$ = 16.8 Hz, $J_{3a,2}$ = 2.6 Hz, Ha-3); ¹³C NMR (DMSO, 100 MHz): δ = 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 ppm; ESI-HRMS calcd for C₁₅H₁₃O₅ [M + H]⁺ 273.0757, found 273.0759.

(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 EtOH to obtain compound **6** as a red solid (0.781 g, 57% yield); mp: 114–115 °C. ¹H NMR (CDCl₃, 400 MHz): δ = 12.29 (brs, 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 ppm (s, 3H, -OCH₃); ¹³C NMR (CDCl₃, 100 MHz): δ = 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 ppm; ESI-HRMS calcd for C₁₈H₁₉O₅ [M + H]⁺ 315.1227, found 315.1232.

2-(3,4-Dimethoxyphenyl)-3-hydroxy-6-methoxy-4H-chromen-4-one (5): An aqueous solution of H₂O₂ (30%, 620 μL) was added to an ice-cold suspension of compound **6** (0.781 g, 2.48 mmol) in EtOH (12 mL) and 1 M NaOH (5 mL). The mixture was allowed to warm to room temperature and was stirred for 4 h. 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 EtOH to obtain compound **5** as pale-yellow powder (0.522 g, 64% yield); mp: 190 °C. ¹H NMR (CDCl₃, 400 MHz): δ = 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 (brs, 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 ppm (s, 3H, -OCH₃); ¹³C NMR (CDCl₃, 100 MHz): δ = 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 ppm; ESI-HRMS calcd for C₁₈H₁₇O₆ [M + H]⁺ 329.1020, found 329.1024.

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 MeOH (10 mL) was added. The reaction mixture was concentrated in vacuo. Water (10 mL) was added, the reaction was concentrated for 1 h 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). ¹H NMR (DMSO, 400 MHz): δ = 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 ppm (d, 1H, J = 8.5 Hz, H-5'); ¹³C NMR

(DMSO, 100 MHz): δ = 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 ppm; ESI-HRMS calcd for C₁₅H₁₁O₆ [M + H]⁺ 287.0550, found 287.0552.

Acknowledgements

We thank Molecular Discovery Ltd. for supplying BioGPS and FLAPdock software for the virtual screening and docking analyses. We also acknowledge AIRC 2015, grants IG10474 and IG16977 given to M.P.C. We thank Dr. Simon Cross for critically reading and revising the manuscript.

Keywords: drug discovery · drug repurposing · virtual screening · thymidylate synthase · cross docking

- [1] D. Farina, F. Spyraakis, A. Venturelli, S. Cross, D. Tondi, M. P. Costi, *Curr. Med. Chem.* **2014**, *21*, 1405–1434.
- [2] F. Spyraakis, *Curr. Drug Targets* **2015**, *17*, 972–973.
- [3] D. Tondi, S. Cross, A. Venturelli, M. P. Costi, G. Cruciani, F. Spyraakis, *Curr. Drug Targets* **2015**, *17*, 983–1005.
- [4] C. Holohan, S. Van Schaeybroeck, D. B. Longley, P. G. Johnston, *Nat. Rev. Cancer* **2013**, *13*, 714–726.
- [5] B. Witkop, *Proc. Am. Philos. Soc.* **1999**, *143*, 540–557.
- [6] A. L. Hopkins, *Nature* **2009**, *462*, 167–168.
- [7] G. V. Paolini, R. H. Shapland, W. P. van Hoorn, J. S. Mason, A. L. Hopkins, *Nat. Biotechnol.* **2006**, *24*, 805–815.
- [8] M. A. Yildirim, K. I. Goh, M. E. Cusick, A. L. Barabasi, M. Vidal, *Nat. Biotechnol.* **2007**, *25*, 1119–1126.
- [9] A. L. Hopkins, *Nat. Chem. Biol.* **2008**, *4*, 682–690.
- [10] T. I. Oprea, J. Mestres, *AAPS J.* **2012**, *14*, 759–763.
- [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*, 134–145.
- [12] A. S. Reddy, S. Zhang, *Expert Rev. Clin. Pharmacol.* **2013**, *6*, 41–47.
- [13] J. Bowes, A. J. Brown, J. Hamon, W. Jarolimek, A. Sridhar, G. Waldron, S. Whitebread, *Nat. Rev. Drug Discovery* **2012**, *11*, 909–922.
- [14] T. T. Ashburn, K. B. Thor, *Nat. Rev. Drug Discovery* **2004**, *3*, 673–683.
- [15] C. R. Chong, D. J. Sullivan, Jr., *Nature* **2007**, *448*, 645–646.
- [16] A. Bairoch, *Nucleic Acids Res.* **1991**, *19*, 2241–2245.
- [17] A. Bairoch, B. Boeckmann, *Nucleic Acids Res.* **1994**, *22*, 3578–3580.
- [18] A. J. Bleasby, D. Akrigg, T. K. Attwood, *Nucleic Acids Res.* **1994**, *22*, 3574–3577.
- [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*, 137–142.
- [21] N. Atias, R. Sharan, *J. Comput. Biol.* **2011**, *18*, 207–218.
- [22] S. Mizutani, E. Pauwels, V. Stoven, S. Goto, Y. Yamanishi, *Bioinformatics* **2012**, *28*, i522–i528.
- [23] E. Pauwels, V. Stoven, Y. Yamanishi, *BMC Bioinf.* **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. Pérez-Nueno, M. Souchet, A. S. Karaboga, D. W. Ritchie, *J. Chem. Inf. Model.* **2015**, *55*, 1804–1823.
- [27] M. Campillos, M. Kuhn, A. C. Gavin, L. J. Jensen, P. Bork, *Science* **2008**, *321*, 263–266.
- [28] J. Scheiber, B. Chen, M. Milik, S. C. Sukuru, A. Bender, D. Mikhailov, S. Whitebread, J. Hamon, K. Azaoui, L. Urban, M. Glick, J. W. Davies, J. L. Jenkins, *J. Chem. Inf. Model.* **2009**, *49*, 308–317.
- [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*, 753–762.
- [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] R. L. Chang, L. Xie, P. E. Bourne, B. O. Palsson, *PLoS Comput. Biol.* **2010**, *6*, e1000938.
- [36] T. Hart, S. Dider, W. Han, H. Xu, Z. Zhao, L. Xie, *Sci. Rep.* **2016**, *6*, 20441.
- [37] S. J. Ho Sui, R. Lo, A. R. Fernandes, M. D. Caulfield, J. A. Lerman, L. Xie, P. E. Bourne, D. L. Baillie, F. S. Brinkman, *Int. J. Antimicrob. Agents* **2012**, *40*, 246–251.
- [38] S. L. Kinnings, L. Xie, K. H. Fung, R. M. Jackson, P. E. Bourne, *PLoS Comput. Biol.* **2010**, *6*, e1000976.
- [39] V. J. Haupt, S. Daminelli, M. Schroeder, *PLoS One* **2013**, *8*, e65894.
- [40] F. Moriaud, S. B. Richard, S. A. Adcock, L. Chanas-Martin, J. S. Surgand, M. B. Jelloul, F. Delfaud, *Briefings Bioinf.* **2011**, *12*, 336–340.
- [41] J. Boström, A. Hogner, S. Schmitt, *J. Med. Chem.* **2006**, *49*, 6716–6725.
- [42] Y. Hu, J. Bajorath, *Drug Discovery Today* **2013**, *18*, 644–650.
- [43] K. Park, S. Lee, H. S. Ahn, D. Kim, *Mol. Biosyst.* **2009**, *5*, 844–853.
- [44] S. Schmitt, D. Kuhn, G. Klebe, *J. Mol. Biol.* **2002**, *323*, 387–406.
- [45] A. Shulman-Peleg, R. Nussinov, H. J. Wolfson, *J. Mol. Biol.* **2004**, *339*, 607–633.
- [46] N. Weill, D. Rognan, *J. Chem. Inf. Model.* **2010**, *50*, 123–135.
- [47] K. Yeturu, N. Chandra, *BMC Bioinf.* **2008**, *9*, 543.
- [48] *SMAP: Software for ligand binding site comparison*, The City University of New York, <http://compsci.hunter.cuny.edu/~leixie/smap/smap.html>.
- [49] T. Liu, R. B. Altman, *PLoS Comput. Biol.* **2011**, *7*, e1002326.
- [50] J. Konc, D. Janezic, *Bioinformatics* **2010**, *26*, 1160–1168.
- [51] L. Siragusa, S. Cross, M. Baroni, L. Goracci, G. Cruciani, *Proteins Struct. Funct. Bioinf.* **2015**, *83*, 517–532.
- [52] L. Siragusa, F. Spyraakis, L. Goracci, S. Cross, G. Cruciani, *Mol. Inf.* **2014**, *33*, 446–453.
- [53] P. J. Goodford, *J. Med. Chem.* **1985**, *28*, 849–857.
- [54] E. Carosati, S. Sciabola, G. Cruciani, *J. Med. Chem.* **2004**, *47*, 5114–5125.
- [55] V. Ferrario, L. Siragusa, C. Ebert, M. Baroni, M. Foscatto, G. Cruciani, L. Gardossi, *PLoS One* **2014**, *9*, e109354.
- [56] M. Baroni, G. Cruciani, S. Sciabola, F. Perruccio, J. S. Mason, *J. Chem. Inf. Model.* **2007**, *47*, 279–294.
- [57] C. W. Carreras, D. V. Santi, *Annu. Rev. Biochem.* **1995**, *64*, 721–762.
- [58] E. Chu, M. A. Callender, M. P. Farrell, J. C. Schmitz, *Cancer Chemother. Pharmacol.* **2003**, *52*, 80–89.
- [59] 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*, 1273–1283.
- [60] 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*, 10272–10276.
- [61] 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*, 5958–5968.
- [62] D. Tondi, A. Venturelli, S. Ferrari, S. Ghelli, M. P. Costi, *J. Med. Chem.* **2005**, *48*, 913–916.
- [63] J. Phan, S. Koli, W. Minor, R. B. Dunlap, S. H. Berger, L. Lebioda, *Biochemistry* **2001**, *40*, 1897–1902.
- [64] G. Ausiello, P. F. Gherardini, E. Gatti, O. Incani, M. Helmer-Citterich, *BMC Bioinf.* **2009**, *10*, 182.
- [65] S. Pérot, O. Sperandio, M. A. Miteva, A. C. Camproux, B. O. Villoutreix, *Drug Discovery Today* **2010**, *15*, 656–667.
- [66] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, *Genome Res.* **2003**, *13*, 2498–2504.
- [67] Z. Zhang, M. G. Grigorov, *Proteins Struct. Funct. Bioinf.* **2006**, *62*, 470–478.
- [68] G. Cruciani, M. Pastor, W. Guba, *Eur. J. Pharm. Sci.* **2000**, *11*, S29–39.
- [69] F. Spyraakis, 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*, 1501–1511.
- [70] F. Spyraakis, 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.
- [71] W. Xu, J. Lim, C. Y. Goh, J. Y. Suen, Y. Jiang, M. K. Yau, K. C. Wu, L. Liu, D. P. Fairlie, *J. Chem. Inf. Model.* **2015**, *55*, 2079–2084.
- [72] W. Xu, A. J. Lucke, D. P. Fairlie, *J. Mol. Graphics Modell.* **2015**, *57*, 76–88.
- [73] L. A. Pinna, F. Meggio, *Prog. Cell Cycle Res.* **1997**, *77*–97.
- [74] S. Sarno, S. Moro, F. Meggio, G. Zagotto, D. Dal Ben, P. Ghisellini, R. Battistutta, G. Zanotti, L. A. Pinna, *Pharmacol. Ther.* **2002**, *93*, 159–168.
- [75] 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–D1090.
- [76] C. Chichester, D. Digles, R. Siebes, A. Loizou, P. Groth, L. Harland, *Drug Discovery Today* **2015**, *20*, 399–405.
- [77] F. A. Kruger, A. Gaulton, M. Nowotka, J. P. Overington, *Bioinformatics* **2015**, *31*, 776–778.
- [78] D. S. Wishart, C. Knox, A. C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, J. Woolsey, *Nucleic Acids Res.* **2006**, *34*, D668–D672.
- [79] L. Xie, J. Wang, P. E. Bourne, *PLoS Comput. Biol.* **2007**, *3*, e217.
- [80] L. Dellafiora, M. Marchetti, F. Spyraakis, V. Orlandi, B. Campanini, G. Cruciani, P. Cozzini, A. Mozzarelli, *Bioorg. Med. Chem. Lett.* **2015**, *25*, 4297–4303.
- [81] G. Muratore, L. Goracci, B. Mercorelli, A. Foeglein, P. Digard, G. Cruciani, G. Palu, A. Loregian, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 6247–6252.
- [82] F. Spyraakis, P. Benedetti, S. Decherchi, W. Rocchia, A. Cavalli, S. Alcaro, F. Ortuso, M. Baroni, G. Cruciani, *J. Chem. Inf. Model.* **2015**, *55*, 2256–2274.
- [83] L. Goracci, S. Buratta, L. Urbanelli, G. Ferrara, R. Di Guida, C. Emiliani, S. Cross, *Eur. J. Med. Chem.* **2015**, *92*, 49–63.
- [84] L. Goracci, M. Ceccarelli, D. Bonelli, G. Cruciani, *J. Chem. Inf. Model.* **2013**, *53*, 1436–1446.
- [85] R. Navakauskienė, M. Mori, M. Christodoulou, A. Zentelyté, B. Botta, L. Dalla Via, F. Ricci, G. Damia, D. Passarella, C. Zilio, N. Martinet, *MedChemComm* **2016**, *7*, 1245–1255.
- [86] 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*, 533–545.
- [87] G. Muratore, B. Mercorelli, L. Goracci, G. Cruciani, P. Digard, G. Palu, A. Loregian, *Antimicrob. Agents Chemother.* **2012**, *56*, 6009–6013.
- [88] D. Cardinale, G. Guaitoli, D. Tondi, R. Luciani, S. Henrich, O. M. Salo-Ahen, S. Ferrari, G. Marverti, D. Guerrieri, A. Ligabue, C. Frassinetti, C. Pozzi, S. Mangani, D. Fessas, R. Guerrini, G. Pontaneri, R. C. Wade, M. P. Costi, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, E542–9.
- [89] 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*, 7773–7785.
- [90] R. Z. Cer, U. Mudunuri, R. Stephens, F. J. Lebeda, *Nucleic Acids Res.* **2009**, *37*, W441–W445.

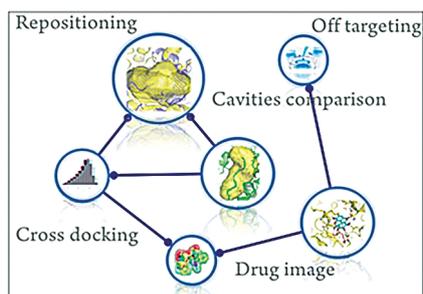
Received: February 29, 2016

Revised: June 8, 2016

Published online on ■■■■■, 0000

FULL PAPERS

The TS test case: We describe the integrated and innovative BioGPS/FLAPdock pipeline for rapid and effective comparison of protein cavities, off-target identification, and drug repurposing. Structural, chemical, and energetic properties of the cavities are simply encoded in the corresponding GRID molecular interaction fields. BioGPS discloses pocket similarity, and cross-docking experiments identify drugs for potential repurposing.



*L. Siragusa, R. Luciani, C. Borsari, S. Ferrari, M. P. Costi, G. Cruciani, F. Spyraakis**

■■ - ■■

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

