ABSTRACT

Motivation: The need for new drugs and new targets is particularly compelling in an era that is witnessing an alarming increase of drug resistance in human pathogens. The identification of new targets of known drugs is a promising approach, which has proven successful in several cases. Here, we describe a database that includes information on 5153 putative drug–target pairs for 150 human pathogens derived from available drug–target crystallographic complexes.

Availability and implementation: The TiPs database is freely available at http://biocomputing.it/tips.

Contact: anna.tramontano@uniroma1.it or allegra.via@uniroma1.it

Received on April 8, 2013; revised on May 8, 2013; accepted on May 16, 2013

1 INTRODUCTION

Novel mechanisms to escape therapy are constantly emerging among human pathogen populations, and this clearly urges the development, on one hand, of new drugs for the treatment of the diseases and, on the other hand, of rapid and effective methods to help expand the landscape of available treatment options (Hopkins et al., 2011). In this context, computational studies are called on to help identify novel therapeutic targets and characterize their interactions, and indeed a number of such efforts are described in the literature (Aguero et al., 2008; Kinnings et al., 2010; Lepore et al., 2011; Orti et al., 2009). However, these are mostly devoted to the analysis of single targets or specific tropical disease pathogens.

The TiPs database has been developed with the aim of facilitating the identification of new therapeutic targets in >150 organisms responsible for human infections. We performed a large-scale analysis to systematically identify candidate targets in the proteomes of such organisms. The rationale of our approach is based on the intrinsic polypharmacological behaviour of compounds targeting homologous proteins (Paolini et al., 2006). We considered all drug–target pairs for which the 3D structure of the complex is experimentally known and used the sequence of the target to identify its homologues in human pathogens. The evolutionary conservation of such homologues and their 3D structures (available or predicted) were used to verify whether the original drug was in principle able to bind them as it does the original target. To this aim, stringent filters were applied to ensure that predicted binding sites and their interactions with the drug are as accurate as possible. Pathogen proteins predicted with high confidence to be therapeutic targets and the putative drugs interacting with them were collected and annotated in TiPs.

2 METHODS

More than 400 human pathogen species were obtained from ‘The Approved List of Biological Agents’ provided by the Advisory Committee on Dangerous Pathogens. To unambiguously assign an identifier (ID) to human pathogens, the names of the organisms were mapped onto the NCBI Taxonomy Database records (http://www.ncbi.nlm.nih.gov/Taxonomy/).

Drug compounds and information on their molecular targets were obtained from DrugBank (http://www.drugbank.ca). The SMILE IDs of drugs annotated either as ‘inhibitor’, ‘agonist’ or ‘antagonist’ were used to associate them with ligands present in the PDB structure entries (Berman et al., 2009). Only identical compounds were considered (Tanimoto coefficient = 1). A total of 308 distinct drugs were observed in complex with at least one PDB structure. About 40% of these (119/308) occur in complex with their actual pharmaceutical target. These were used as starting points to predict potential drug targets in pathogens. The search for homologues in pathogens was performed using BLAST+ (Camacho et al., 2009) with default parameters against the nr database (ftp://ftp.ncbi.nlm.nih.gov/blast/db/). We only retained highly reliable hits, i.e. those showing at least 40% sequence identity to the original target and e-value < 10^{-6}.

Pathogen taxonomic IDs were retrieved by matching the gi numbers of predicted pathogen sequences were retrieved through a multiple sequence alignment from any atom of the drug. The drug-binding site residues in the subset of target residues having at least one PDB structure. About 40% of these (119/308) occur in complex with their actual pharmaceutical target. These were used as starting points to predict potential drug targets in pathogens. The search for homologues in pathogens was performed using BLAST+ (Camacho et al., 2009) with default parameters against the nr database (ftp://ftp.ncbi.nlm.nih.gov/blast/db/). We only retained highly reliable hits, i.e. those showing at least 40% sequence identity to the original target and e-value < 10^{-6}.

Pathogen taxonomic IDs were retrieved by matching the gi numbers of predicted pathogen sequences were retrieved through a multiple sequence alignment from any atom of the drug. The drug-binding site residues in the subset of target residues having at least one PDB structure. About 40% of these (119/308) occur in complex with their actual pharmaceutical target. These were used as starting points to predict potential drug targets in pathogens. The search for homologues in pathogens was performed using BLAST+ (Camacho et al., 2009) with default parameters against the nr database (ftp://ftp.ncbi.nlm.nih.gov/blast/db/). We only retained highly reliable hits, i.e. those showing at least 40% sequence identity to the original target and e-value < 10^{-6}.

Pathogen taxonomic IDs were retrieved by matching the gi numbers of predicted pathogen sequences were retrieved through a multiple sequence alignment from any atom of the drug. The drug-binding site residues in the subset of target residues having at least one PDB structure. About 40% of these (119/308) occur in complex with their actual pharmaceutical target. These were used as starting points to predict potential drug targets in pathogens. The search for homologues in pathogens was performed using BLAST+ (Camacho et al., 2009) with default parameters against the nr database (ftp://ftp.ncbi.nlm.nih.gov/blast/db/). We only retained highly reliable hits, i.e. those showing at least 40% sequence identity to the original target and e-value < 10^{-6}.
generated using the Modeller software. Note that the best template used to build the models corresponds to the original structure in the druggable complexes only in 153 cases, whereas in all the other cases, the best template was a different structure.

The binding site residues of the original complex and of the predicted target were structurally superimposed using the LGA software (Zemla, 2003). Subsequently, the ligands were transferred into the structure or model of the pathogen protein that could be successfully superimposed within 5 Å distance to the known target. Binding sites in the modeller structures were analyzed for the occurrence of nearby insertions/deletions. These cases are suitably highlighted in the TiPs database search output. This allows users to analyse them to establish the likelihood that their presence affects the conformation of the binding site.

3 RESULTS

TiPs currently contains 4071 candidate pathogen target structures involved in 5153 different drug-target complexes in 150 pathogens. All entries are thoroughly annotated with both sequence and functional information. The database can be queried by organism name (genus or species name), protein family or function (EC number, GO terms, P:fam), as well as UniProt ID. The query returns a sortable table providing information about both known and predicted drug-target pairs and links to visualize specific information on the drug(s) (physicochemical properties, structure, indication and side effects), the target(s) (UniProt annotation and PDB structure(s)) and to visually analyse or download their 3D complexes. LigPlot (Laskowski and Swindells, 2011) drawings of both the known and inferred binding sites in complex with the drug are available as well (Fig. 1).

ACKNOWLEDGEMENT

The authors are grateful to all members of the group for useful suggestions.

Funding: This work was supported by the King Abdullah University of Science and Technology (KAUST), Award No. KUK-I1-012-43, PRIN 20108XYHJS and FIRB RBIN06E9Z8_005.

Conflict of Interest: none declared.

REFERENCES

Aguero, F. et al. (2008) Genomic-scale prioritization of drug targets: the TDR targets database. Nat. Rev. Drug Discov., 7, 900–907.

Berman, H.M. et al. (2012) The protein data bank at 40: reflecting on the past to prepare for the future. Structure, 20, 391–396.

Camacho, C. et al. (2009) BLAST+: architecture and applications. BMC Bioinformatics, 10, 421.

Hopkins, A.L. et al. (2011) Rapid analysis of pharmacology for infectious diseases. Curr. Top. Med. Chem., 11, 1292–1300.

Kinnings, S.L. et al. (2010) The Mycobacterium tuberculosis drugome and its polypharmacological implications. PLoS Comput. Biol., 6, e1000976.

Kopp, J. and Schwede, T. (2004) Automated protein structure homology modeling: a progress report. Pharmacogenomics, 5, 405–416.

Laskowski, R.A. and Swindells, M.B. (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J. Chem. Inf. Model., 51, 2778–2786.

Lepore, R. et al. (2011) Identification of the Schistosoma mansoni molecular target for the antimalarial drug artemether. J. Chem. Inf. Model., 51, 3005–3016.

Orti, L. et al. (2008) A kernel for open source drug discovery in tropical diseases. PLoS Negl. Trop. Dis., 3, e418.

Paolini, G.V. et al. (2006) Global mapping of pharmacological space. Nat. Biotechnol., 24, 805–815.

Remmert, M. et al. (2012) HHBlits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat. Methods, 9, 173–175.

Taly, J.F. et al. (2011) Using the T-Coffee package to build multiple sequence alignments of protein, RNA, DNA sequences and 3D structures. Nat. Protoc., 6, 1669–1682.

Zemla, A. (2003) LGA: a method for finding 3D similarities in protein structures. Nucleic Acids Res., 31, 3370–3374.