Summary

Sirtuins are evolutionarily conserved NAD^+ dependent deacetylase enzymes that belong to class III histone deacetylase family [28]. Initially, sirtuins were implicated in the process of gene silencing and longevity [14, 15] but has since been shown to participate in many other biological functions [16]. Sirtuins regulate a number of cellular processes including cell cycle, apoptosis, metabolic regulation, ageing and inflammation. The growing evidence in last few years has established their involvement in various diseases such as type II diabetes mellitus, cancer, ageing, Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis and cardiovascular diseases [31, 37, 39, 46, 84, 86-89, 97]. Recent past has witnessed discovery of substantial number of sirtuin modulators. However, many of these modulators are either non-specific or their specificity is not fully established. While different sirtuins and histone deacetylase classes perform distinct cellular processes, their specific functions are not absolutely defined. In many cases, more than one sirtuin are found to deacetylate same acetylated protein targets, for e.g. histone H4 at lysine 16 is deacetylated by both SIRT1 and SIRT2 [155, 156]. This is due to the fact that all sirtuins share a highly conserved catalytic core. However, the overlapping functional and substrate features of different sirtuins can prove to be useful in design and development of small molecule therapeutics for different diseases. The less selective inhibitors may find their application where highly specific inhibitors may not be able to fully manipulate the specific cellular roles of a given sirtuin. For instance, cancer treatment may involve broad spectrum inhibitors that target different sirtuins or HDACs. On the contrary, the highly selective inhibitors may prove to be effective to reduce the side effects that arise due to perturbed function of a particular sirtuin isoform. Therefore, it would be beneficial to develop both highly selective and broadly active therapeutic agents for sirtuins. Further, considering the fact that in past majority of small molecule inhibitors identified target SIRT1, it is important to develop small molecule therapeutics for other SIRT members. Development of sirtuin activators is also important for the treatment of various diseases.

Owing to their promising role in pharmaceutical intervention, sirtuins have captured the attention of researchers to develop potential drugs. However, drug development is a complex and time-consuming process challenged by high rates of failure due to factors such as problematic functionalities, solubility at relevant concentrations, toxicity and off-target effects [218, 219]. It is critical to design a focused library of molecules in order to avoid possible risk factors in later stages of drug discovery. In this context, computational methods can rationally guide the search for active compounds by utilizing the accumulated information in literature. We developed a database called EpiDBase (www.epidbase.org) to explore, search and analyze modulators for sirtuins as well as other epigenetic proteins [176]. Epigenetic proteins are responsible for many crucial cellular processes and any disruption in their activity leads to different pathological conditions such as cardiovascular disease, metabolic disorders and cancer. Despite potentially huge role of epigenetic targets in human diseases, it is still underexplored area. EpiDBase analyzes and presents unique set of drug like ligands, shows their ADMET profile, allows text or structure derived search and supports chemoinformatic analysis would be highly beneficial for drug discovery scientists. EpiDBase is also beneficial to analyze the chemical space and scaffolds of sirtuin modulators, in order to design and acquire a focused library of compounds for screening.

The concept of chemical scaffold diversity is widely applied in medicinal chemistry. The chemical space representing just drug like molecules (*i.e.* between 300 to 500 Da of molecular weight) is estimated to encompass 10^{60} molecules [186]. Practically, it is impossible to mine out drug molecules from this vast chemical space. Thus, in drug discovery it is important to select a chemical library with overlapping biological and chemical space [187]. The scaffolds should be analyzed as the core of chemical libraries prior to synthesis and/or acquisition. In EpiDBase, we collected data pertaining to epigenetic proteins, their ligands and relevant information from the literature. The complex structures such as peptide molecules and macrocycles were removed from the collected ligands. In order to retain only drug-like molecules, ligands were passed

through ZINC filter and Lily MedChem rules. Open Babel was used to analyze the physicochemical properties and ligands were clustered using ChemMine tools.

EpiDBase has a total of 11,422 entries encompassing 5,748 unique ligands associated with 220 different epigenetic proteins. The database can be accessed to retrieve the information pertaining to epigenetic proteins, their reported ligands, experimental IC_{50} values, structural data, toxicological profile and chemoinformatic data. The database can be browsed using protein family names, ligand structures, substructures and fingerprint based chemical similarity search. EpiDBase can be utilized for structure activity relationship studies, statistical analysis, fragment based drug design, virtual screening and molecular docking studies to assist epigenetic drug discovery. In general, this knowledge base can be useful to design and discover modulators to influence epigenetic states of various diseases including cancer, diabetes, neurodegenerative disorders and cardiovascular disorders. The sirtuin modulators were clustered using a Tanimoto coefficient cut-off of 0.6. Also, a conformer library was generated using modulators of sirtuins.

In epigenetic drug discovery, the systematic elimination of PAINS can prevent the wastage of resources and time. For this purpose, we collected the ligands from major epigenetic databases including EpiDBase, ChEMBL and PubMed, to analyze them for the occurrence of PAINS and drug-like properties of epigenetic modulators. We found a total of 65,039 PAINS out of 4,54,044 epigenetic ligands and eliminated them to retain a curated library of 3,89,005 molecules. The clustering analysis of PAINS unveiled the most frequent scaffolds in epigenetic drug discovery that included rhodanines, phenyl-sulfonamides, isothiozolones and quinones. After PAIN elimination, the resulting curated library was subjected to chemoinformatic ADMET analysis to predict druglikeness. On average, the ADMET properties of ligands of various epigenetic proteins including DNMT3A, DNMT3B, DOT1L, EHMT2, PCAF, PRMT5, SETD7, SMYD2, SMYD3, CBX1, CBX7, KDM4C, KDM5A, KDM6A, KDM6B, KDM7A, SIRT4, SIRT5, SIRT6 and SIRT7 were significantly higher than the drug-like score. The curated library was

also subjected for hit prioritization. TL score was assigned to epigenetic ligands of each protein by using the mean values of four descriptors including MW, clogP, TPSA and nrotB. On average, the ligands of DNMT3A, PCAF, TIP60, SUV420H2, SUV420H1, SUV39H1, SMYD3, SMYD2, SETD7, PRDM1, PRDM10, PRDM11, KDM4C, SIRT4, SIRT5, SIRT6, SIRT7, CBX1, CBX4, CBX7 and CBX8 showed higher values of TL score indicating poor possibility of becoming a drug. Our comprehensive chemoinformatic analysis has demonstrated that many of known epigenetic ligands lack drug-likeness and therefore it is essential to synthesize new molecules with lead-like properties for successful epigenetic drug discovery.

We have designed a focused library of small molecules and fragments utilizing the EpiDBase for conducting the screening assays for sirtuins. To perform the in vitro screening, purified SIRT proteins were required. Thereby, clones of human sirtuins (SIRT1 to SIRT7) were generated, transformed and expressed in *E. coli*. Recombinant expression and purification of SIRT1, SIRT2, SIRT3-truncate, SIRT5, SIRT6 and SIRT7 was achieved successfully. Label-free peptides were synthesized to test the activity of sirtuins. The p53K15ac peptide showed activity for SIRT2, SIRT3 and SIRT5. The MALDI-TOF based assay was optimized to screen the focused library against SIRT2, SIRT3 and SIRT5. Screening identified a total of seven inhibitors against SIRT2, SIRT3 and SIRT5. Compound 3 and Compound 4 were found to inhibit all SIRT2, SIRT3 and SIRT5. Compound 15 was found inhibiting both SIRT3 and SIRT5. Compounds 8, compound 12, and compound 18 were specific to SIRT3. The ratio-metric quantification was performed to determine the IC_{50} value of theses inhibitors. The IC_{50} of compound 3 and compound 4 was in low micromolar range for each SIRT2, SIRT3 and SIRT5. The specific inhibitor of SIRT3, compound 8 had IC_{50} of 78.12 μM. The docking analysis revealed that hit molecules identified in this study bind in the catalytic site of SIRT2, SIRT3 and SIRT5. Compound 3 and compound 4 show interactions with the residues at NAD^+ binding pocket B and C of SIRT2. All hits identified against SIRT3 binds in the NAD^+ binding pocket except for compound 3 and compound 18, which show interactions with substrate binding site as well. Hit molecules found against SIRT5 also bind in the

NAD^+ binding site, additionally Compound 4 interacts with residues from substrate binding site. Crystallization trials of SIRT2 and SIRT3 were attempted to characterize the interactions between hits and protein. Various commercial crystallization screens and conditions were applied, but were not successful. PNC1-OPT coupled assay was optimized to screen a library of fragments against SIRT5. The fragment screening for SIRT5 using PNC1-OPT assay identified some novel fragments. The FR1 fragment showed significant increase in florescence signal indicating enhanced SIRT5 activity, whereas fragments FR22, FR31 and FR47 showed decrease in SIRT5 activity.