Lior's Thesis

Results

Description of the dataset

The first step in training our classifier was obtaining a dataset of peptides that had their activity level with the respect to the HDAC8 enzyme verified experimentally. *Fierke et al* created a dataset composed of 361 6-mer peptides with the sequence GXK(Ac)YGC (where X,Y are all the amino acids except Cysteine). For each of these peptides, a level of activity with respect to HDAC8 was determined by measuring the percentage of deacetylation after 1 hour.(?) (**Add reference to the proper section in the supplementary material**) The dataset was divided to training and test sets by sorting the peptides by their activity , taking all the even rows to be the test set and all the odd rows to be the training set. That division assured even distribution of peptides with respect to their activity levels (avoiding a situation where one set holds a large number of high/low activity decoys).

Template selection

As we've previously discussed, our protocol models the interaction between a peptide and its corresponding receptor. FlexPepDock takes as input a three dimensional structure of the receptor and a low resolution approximation of the peptide. In our case, the receptor is HDAC8, its three dimensional structure was solved on numerous occasions and under different conditions in the last few years. In this study we tried to use multiple structures as our template, hoping that one of them will give an accurate complex with the peptidic substrates. Below is a table that summarizes the structures that were tested as templates for this study:

PDB ID	Reference	Description
2v5w	[1]	HDAC8 in complex with a p53-derived diacetylated peptide with a Y306F catalysis abolishing mutation
3f07	[2]	HDAC8 complexed with APHA
3ew8	[3]	HDAC8 solved as a monomer, with a catalysis abolished mutation: D101L
1t67	[4]	HDAC8 complexed with hydroxamate inhibitor (MS-344), residues 62-68 were discarded from the model

Choosing the right template is a formidable challenge - some structures were solved with inhibitors - a thing that could induce a different *bound* structure than the actual real substrates. Others were solved with mutations that abolished catalysis and/or binding. And most of all, most structures were solved as dimers that interacted with their highly flexible regions, creating crystal contacts and potential interactions that might have altered the specificity profile of the enzyme.

Preparation of starting structure

For each of the peptide sequences a coarse model of the complex was generated based on the selected template, that coarse model is the starting structure that serves as the input to the FlexPepDock protocol. According to the *no free lunch theorem*, all search algorithms have the same average performance over all problems, and thus implies that to gain in performance on a certain application one must use a specialized algorithm that includes some prior knowledge about that problem. In previous studies we found that incorporating key interactions between the peptide and the receptor as constraints in FlexPepDock's search algorithm greatly improves the performance of the resulting predictor.

Unlike previous studies, where the key interactions from which the constraints were derived relied heavily on backbone atoms, we found that the dominant interactions in our case are mostly concentrated around the acetylated Lysine.

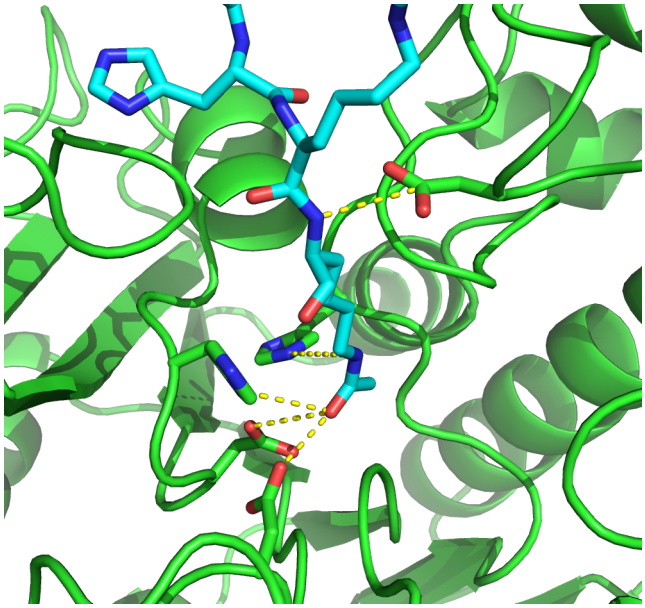

Figure 1: *The key interactions from which the constraints were derived, taken from a solved crystal complex (PDB: 2v5w).*

1. taking the backbone from the native peptide of 2v5w, fixing the acetylated lysine , extending and mutating the residues to the desired peptide sequence.

Calibration of the protocol

1. Explanation of scoring terms
2. **Different approaches for modeling (10 representatives)**
    1. Naive scoring mechanism
    2. approach with different scoring
    3. Clustering and how it influenced the results

Whole data set analysis

1. measures of success
2. determination of cutoff
3. statistical tests

Phosphosite database

1 Vannini A, Volpari C, Gallinari P, et al. Substrate binding to histone deacetylases as shown by the crystal structure of the HDAC8-substrate complex. EMBO Rep. 2007;8(9):879-84.
2 Dowling DP, Gantt SL, Gattis SG, Fierke CA, Christianson DW. Structural studies of human histone deacetylase 8 and its site-specific variants complexed with substrate and inhibitors. Biochemistry. 2008;47(51):13554-63.
3 Dowling DP, Gantt SL, Gattis SG, Fierke CA, Christianson DW. Structural studies of human histone deacetylase 8 and its site-specific variants complexed with substrate and inhibitors. Biochemistry. 2008;47(51):13554-63.
4 Somoza JR, Skene RJ, Katz BA, et al. Structural snapshots of human HDAC8 provide insights into the class I histone deacetylases. Structure. 2004;12(7):1325-34.