

# Applying Sequential Forward Floating Selection to Protein Structure Prediction with a Study of HIV-1 RP

Jeff Reneker, Chi-Ren Shyu, Ph.D.

Department of Computer Science  
Department of Health Management & Informatics  
University of Missouri – Columbia  
Columbia, MO 65211

*Accurate structural models of target proteins are an essential component to structure-based drug design methods [1, 2]. To aid in this effort, the sequential forward floating selection (SFFS) algorithm [3] is being applied to protein structure prediction. This algorithm demonstrates a reasonable balance between optimality of feature selection and efficiency while using large datasets such as the protein databank (PDB) [4]. The resulting protein structure predictions will be used in a study of HIV-1 PR.*

## ALGORITHM OVERVIEW AND APPLICATION

We are currently specializing the general-purpose SFFS algorithm to apply to protein structure prediction. The input to our algorithm is an amino acid sequence of a test protein and the output is short, ranked list of candidate structures. The three steps to the SFFS algorithm are outlined as follows. Briefly, the first step (inclusion) identifies the most significant feature among unselected features with respect to the current set of selected features. In our context, features are PDB subsequences and corresponding substructures. The second step (exclusion) identifies the least significant feature in the current feature set and removes it unless it is the feature previously added in step 1. The third step is a continuation of the feature exclusion process that is based on the significance of the least significant feature in the current feature set as compared to previous feature sets of the same cardinality. The resulting set of features (substructures) is a close-to-optimal set based on the initial set from the PDB. These substructures are rotated and linked together to form candidate structures representing an amalgam of structural information from the PDB. Accurate structures can be generated by carefully choosing the initial set. Additionally, the algorithm works within a reasonable period of time.

Our protein structure predictions will be used in a study of the retroviral protease, HIV-1 PR. This protease is the target of several different drugs including Saquinavir, Ritonavir, Indinavir, and Nelfinavir. Since HIV RT is unable to edit

transcription errors during nucleic acid replication, HIV-1 PR mutates rapidly and can become resistant to current drugs. For this reason, there is a continual need to develop new drugs, often using structure-based methods. A study of clinical isolates of HIV-1 PR [5] revealed five non-contiguous, conserved regions of the protein. These five regions will serve as benchmarks for structure predictions returned from our SFFS algorithm. By varying the sequence of HIV-1 PR and examining our resulting structure predictions, we can rapidly study which changes potentially alter the structure the least (and the most) and, therefore, which amino acids are favorable (or unfavorable) targets for new drugs.

## ACKNOWLEDGEMENT

JR is funded from National Library of Medicine (NLM) Bioinformatics and Health Informatics Training (BHIRT) Grant No. 5 T15 LM07089 13 and CRS is supported by the Shumaker Endowment in Bioinformatics.

## REFERENCES

1. Christine S. Ring, Eugene Sun, James H. McKerrow, Garson K. Lee, Philip J. Rosenthal, Irwin D. Kuntz, and Fred E. Cohen (1993). Structure-based inhibitor design by using protein models for the development of antiparasitic agents. *Proc. Natl. Acad. Sci. USA*, **90**, 3583-3587.
2. Alexander Wlodawer and Jiri Vondrasek (1998). Inhibitors of HIV-1 protease: A major success of structure-assisted drug design. *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 249-284.
3. P. Pudil, J. Novovicova, and J. Kittler (1994). Floating search methods in feature selection. *Pattern Recognition Letters*, **15**, 1119-1125.
4. <http://www.rcsb.org/pdb/Welcome.do>
5. D.L. Winsow, S. Stack, R. King, H. Scarnati, A. Bincsik, and M.J. Otto (1995). Limited sequence diversity of the HIV type 1 protease gene from clinical isolates and in vitro susceptibility to HIV protease inhibitors. *AIDS Res. Hum. Retrovir.*, **11**, 107-113.