Mapping the Structure Space of the Ras Protein using a Novel Hybrid Evolutionary Algorithm

Rudy Clausen[1,2], Dr. Amarda Shehu[1], Dr. Buyong Ma[2], and Dr. Ruth Nussinov[2]

[1]Department of Computer Science, George Mason University, Fairfax VA.

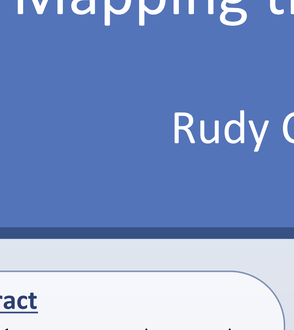
[2]National Cancer Institute, National Institute of Health, Frederick MD.

Abstract

While many proteinopathies are caused by sequence mutations removing the ability of a protein to assume a specific structure, some of the most complex human diseases are not so easily explained. Mutations may not invalidate structures populated by the wildtype protein but instead affect the rate at which the protein switches between structures. This emerging picture of proteins as dynamic systems switching between structures to modulate function must be taken into account when studying protein structure in disease and therefore demands a comprehensive structural characterization only possible through an energy landscape treatment. Only sample-based representations of a protein energy landscape are viable in silico, and sampling-based exploration algorithms have to address the fundamental but challenging issue of balancing between exploration (broad view) and exploitation (going deep). We propose here a novel algorithm that combines concepts from evolutionary computation and protein modeling research to achieve this balance. The algorithm leverages experimental structures by using principal component analysis to obtain a reduced space from which the algorithm can draw samples. Samples are then lifted from the reduced to an all-atom structure space where they are then mapped to nearby local minima in the all-atom energy landscape. The proposed algorithm is used to make the first steps towards answering the question of how sequence mutations affect the function of the Ras protein by providing the energy landscape as the intermediate explanatory link between protein sequence and function.

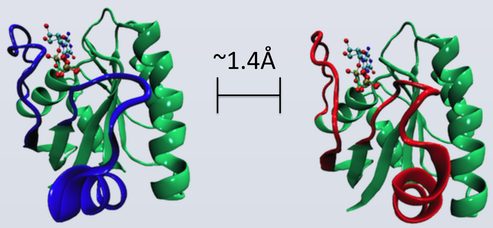
Ras Protein

- Mediates signaling pathways that control cell proliferation and growth.
- Binds GTP/GDP, where GTP puts Ras in an active state, and GDP puts Ras in an inactive state.
- Mutations that affect this GTP/GDP cycle can lead to cancer.
  - Mutations that affect this GTP/GDP cycle are found in over 20% of all human cancers. (1)
- Many mutations have been studied due to the involvement of Ras in cancer.
  - Point mutations such as G12V can cause Ras to be stuck in an active state (2).
  - Mutations may also cause Ras to activate at inappropriate times.
- Many structures of Ras that contain mutations have been solved and deposited in the PDB.



GDP-bound (red) and GTP-bound(blue)

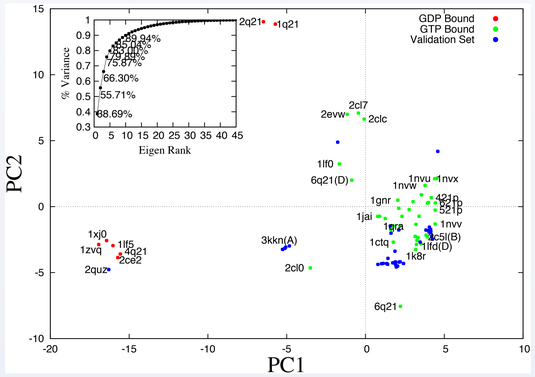
Conformational Changes of Interest



GDP Bound GTP Bound

- Conformational Change between the two functional states is very small (~1.4Å).
- Need to explore high dimensional structure space of Ras protein (498D) while balancing exploration (broad) and exploitation (deep)

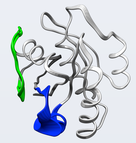
Principal Component Analysis



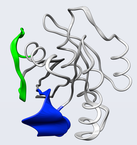
- Use PC Space to reduce search space from 498 dimensions to 10 while preserving 90% of the variance of the original data set.
  - Can now explore this smaller and reduced space instead
- PC1 and PC2 motions correspond to know conformational changes between GTP and GDP bound states
  - Searching this space will focus on known conformational changes
- PC space well separates the 2 known conformations (GTP and GDP bound)

Movement Captured by

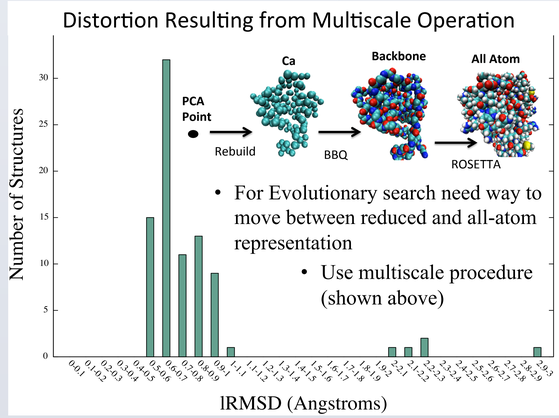
PC1



PC2



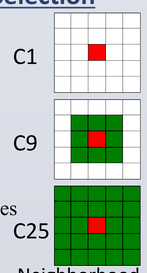
Distortion Resulting from Multiscale Operation



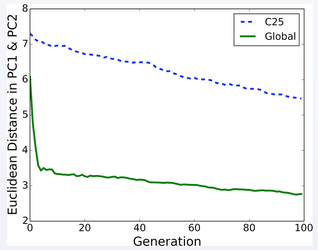
- For Evolutionary search need way to move between reduced and all-atom representation
  - Use multiscale procedure (shown above)

Evolutionary Search using Local Selection

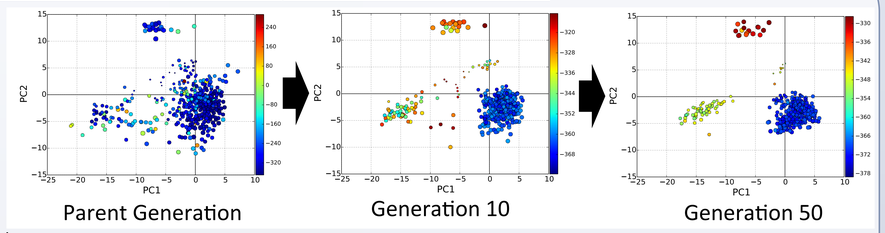
- Evolutionary Search algorithm
  - Population of 500 structures
  - 100 generations
- Use local selection for comparing children to parents
  - Interested in local minima that correspond to potential metastable states
  - Prevents early convergence to global minima



Evolutionary Search using Local Selection (Cont.)

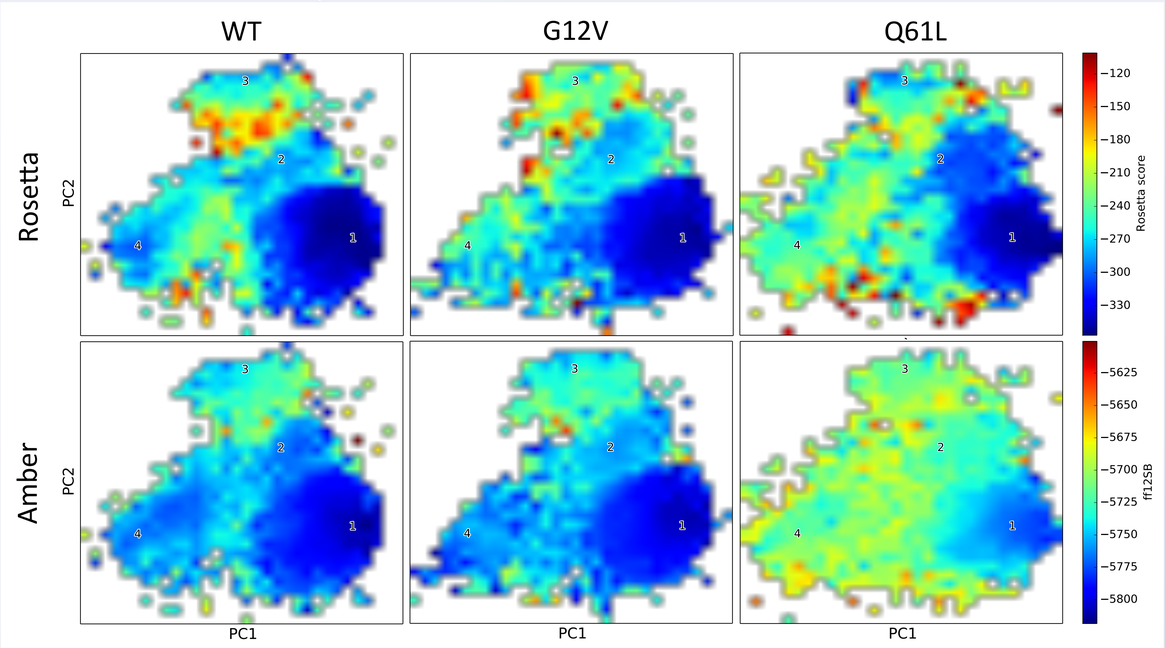


- When a global selection operator is used, the population converges extremely fast
- When a local selection operator (C25 for our experiments) is used, convergence is gradual allowing time to converge to local minima.



- Figure above depicts the gradual convergence as the Evolutionary Search continues
- The population clearly converges to multiple locations within the PC space as opposed to a single global minima
  - Each local minima has a slightly different energy depth
- Average energy of the population either remains unchanged or decreases with every new generation.

Resulting Energy Landscapes



- For each experimental setting (Ras WT, G12V, and Q61L) all structures generated with a Rosetta energy score less than -100 were used to produce an energy landscape
  - The same set of structures were then reevaluated by Amber for which the values were then used to reproduce the energy landscape
  - 4 Energy minima detected in WT landscapes for both Rosetta and Amber
- Amber and Rosetta are mostly in agreement about the resulting energy landscape
  - G12V showed very little change in the energy landscape suggesting the loss of function by this mutation is due to a change in binding affinity
  - Q61L showed a large change in the energy landscape indicating a loss of stability for any structure other than the one corresponding to GTP bound

Acknowledgements

Many of the computations were run on ARGO, a research computing cluster provided by the Office of Research Computing at George Mason University, VA (URL: http://orc.gmu.edu). Funding for this work is provided in part by the National Science Foundation (Grant No. 1421001 and CAREER Award No. 1144106) and the Thomas F. and Kate Miller Jeffress Memorial Trust Award. The authors are also grateful to the members of the Shehu lab for useful feedback during this work.

References

1. Julian Downward, Targeting RAS signalling pathways in cancer therapy, *Nature Reviews Cancer* 3, 11-22, 2003
2. M. Takaoka, H. Harada, T. B. Deramaudt, K. Oyama, C. D. Andl, C. N. Johnstone, B. Rhoades, G. H. Enders, O. G. Opitz, and H. Nakagawa. Ha-Ras(G12V) induces senescence in primary and immortalized human esophageal keratinocytes with p53 dysfunction. Oncogene, 23(40):6760{6768, 2004.