Exploring the Structure Space of Wildtype Ras Guided by Experimental Data

Rudy Clausen
Department of Computer Science,
George Mason University,
Fairfax, VA 22030
rclausen@gmu.edu

Amarda Shehu*
Department of Computer Science,
Department of Bioengineering
School of Systems Biology
George Mason University,
Fairfax, VA 22030
amarda@gmu.edu

ABSTRACT

The Ras enzyme mediates critical signaling pathways in cell proliferation and development by transitioning between GTP- (active) and GDP-bound (inactive) states. Many cancers are linked to specific Ras mutations affecting its conformational switching between active and inactive states. A detailed understanding of the sequence-structure-function space in Ras is missing. In this paper, we provide the first steps towards such an understanding. We conduct a detailed analysis of X-ray structures of wildtype and mutant variants of Ras. We embed the structures onto a low-dimensional structure space by means of Principal Component Analysis (PCA) and show that these structures are energetically feasible for wildtype Ras. We then propose a probabilistic conformational search algorithm to further populate the structure space of wildtype Ras. The algorithm explores a low-dimensional map as guided by the principal components obtained through PCA. Generated conformations are rebuilt in all-atom detail and energetically refined through Rosetta in order to further populate the structure space of wildtype Ras with energetically-feasible structures. Results show that a variety of novel structures are revealed, some of which reproduce experimental structures not subjected to the PCA but withheld for the purpose of validation. This work is a first step towards a comprehensive characterization of the sequence-structure space in Ras, which promises to reveal novel structures not probed in the wet laboratory, suggest new mutations, propose new binding sites, and even elucidate unknown interacting partners of Ras.

Keywords

Ras; structure space; energy surface; stable structures; principal components; probabilistic search; multiscaling

*Corresponding Author

BCB '13, September 22 - 25, 2013, Washington, DC, USA

Categories and Subject Descriptors

G.3 [**Probability and Statistics**]: Probabilistic algorithms; J.3 [**Computer Appplications**]: Life and Medical Sciences

General Terms

Algorithms

1. INTRODUCTION

Ras is an important protein in signaling pathways responsible for cell growth and proliferation [12]. Ras accomplishes this by switching between a GTP-bound active state and a GDP-bound inactive state. Two structures representative of the active and inactive state, respectively, are shown in Figure 1. The structures are superimposed over each-other to show that the main structural difference is localized to a region of the structure associated with the binding site.

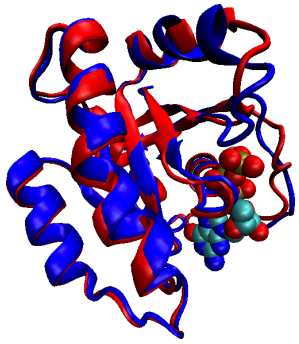


Figure 1: The GDP-bound state, represented by PDB id 4q21, is drawn in red and superimposed over the GTP-bound state represented by PDB id 1qra, which is drawn in blue. GTP is shown in the binding site. Rendering is performed with the Visual Molecular Dynamics (VMD) software [10].

Multiple isoforms are known, but the most studied one is H-Ras, which we refer to as Ras from now on and make the subject of our investigation here. A large range of mutations of Ras have been studied. Mutations that affect the conformational switching in the Ras GTP/GDP cycle often lead to cancer and developmental diseases. In fact, mutations in Ras are found in over 25% of human cancers [5]. Point mutations such as G12V cause Ras to become stuck in the active state and not accomplish the needed conformational

switching, resulting in premature senescence and changes to cellular morphology [23]. Other known mutations cause Ras to switch into the active state at inappropriate times. Due to the wide range of mutations and their role in human cancers, many structures of Ras variants have been solved, primarily through X-Ray Crystallography, and have been deposited in the Protein Data Bank (PDB) [2].

The large number of structures documented for wildtype and mutant variants of Ras can be considered to provide information on the structure space available for biological function. Important questions can be posed. For instance, why is it that Ras is prone to so many mutations that instead of having a deleterious effect on folding end up in energetically-accessible structures affecting conformational switching? The obvious hypothesis is that the Ras energy surface is rather flat, allowing this enzyme to hop among different stable structures upon rather few mutations.

We investigate this hypothesis here. First, a Principal Component Analysis (PCA) is conducted to obtain a low-dimensional embedding of the space of crystal structures. Crystal structures of wildtype and mutant variants are reconstructed from few principal components (PCs), threaded onto the wildtype sequence, and energetically refined to obtain a low-dimensional potential energy map for wildtype Ras. The map shows that many different structures are indeed of low energy, supporting the hypothesis that the energy surface of wildtype Ras is flat and particularly prone to mutations that access different equilibrium structures.

The information available from the PCA is used to further explore the structure space and reveal novel low-energy structures not currently probed in the wet laboratory. The algorithm is guided by a few PCs revealed by the PCA to capture the greatest variation among crystal structures. While the sampling of novel structures is conducted at C_α level detail, each sampled trace is added all-atom detail and energetically refined through the Rosetta all-atom relaxation protocol [13]. Structures are sampled by growing a tree in the structure space, extending paths until encountering high-energy structures, at which point sampling begins anew from any sampled structure. The analysis of sampled structures suggests that many of them reproduce crystal structures withheld from the PCA for the purpose of validation; the algorithm is shown to populate new energetically-feasible regions of the wildtype Ras structure space.

It is worth noting that the investigation presented in this paper is the first to propose exploration of the structure space of Ras that is not based on Molecular Dynamics (MD). In general, exploration of the structure space relevant for biological function is computationally challenging and can currently only be addressed by probabilistic search and optimization techniques [21,22]. In particular, approaches based on Molecular Dynamics (MD) have limited exploration capability. MD-based investigations of Ras focus on modeling fluctuations around a specific structural state or capturing conformational switching between states [3,4,6,7]. However, the most recent MD study on Ras shows that it is rather hard to obtain structures that bridge the structure space in between the active and inactive states [7].

The algorithm proposed in this paper takes a complementary approach that is not based on Molecular Dynamics. Three key components contribute to its computational efficiency. First, a tree-based exploration is carried out in the structure space of Ras in contrast with single-trajectory explorations in MD-based approaches. Second, the search is carried out in a low-dimensional space of PCs that is extracted from available experimental structures of Ras in the PDB. In fact, the extracted information is used to guide the tree-based exploration. Third, the algorithm employs multiscaling, gradually increasing the level of representational detail in order to efficiently obtain all-atom low-energy structures of Ras. The results presented here show that the algorithm has high sampling capability.

It is worth noting that the proposed algorithm bears some resemblance with robotics-inspired graph- or tree-based conformational search algorithms presented to sample the native state of a protein sequence [18–20], compute conformational paths connecting diverse functionally-relevant states of the same sequence [9, 11, 16], or obtain folding pathways of a protein given its native structure [17, 24]. However, the only resemblance of consequence is the employment of a graph-like search structure, as other methods rely on specific representations and sampling strategies closely tied to the application at hand. As such, these methods are not easily portable to mapping the functionally-relevant structure space of a protein. The work presented in [1, 14] to explore peptide and protein conformational spaces is most closely related to the algorithm proposed here. The guidance by a few PCs here is similar to how Normal Modes (NMs) are used in [1, 14]. However, while the NMs are extracted from an analysis of a single structure, the work here uses PCs extracted from analysis of an entire set of X-ray structures. Moreover, the employment of different resolutions and the strategy used to grow the tree further in the structure space are key novel components in the proposed algorithm.

The work provided in this paper is a first step towards a comprehensive map of the sequence-structure space of Ras. In the preliminary investigation conducted here, we focus on the wildtype sequence. However, the presented analysis and exploration can be repeated for any selected variant. Since the PCA is conducted over C_α traces of crystal structures, these structures can be threaded back onto any selected sequence, not only wildtype, to show energetic variations of the structure space depending on the sequence of focus. The algorithm can then further populate the structure space for the selected sequence. Work in this direction will allow determining, for instance, how mutations might affect the relative stability of equilibrium structures and possibly lead to determining novel interacting partners and formulating hypotheses on the role of mutations in certain cancers.

2. METHODS

2.1 Low-dimensional Embedding of Space of Crystal Structures

All 105 experimentally-available structures of Ras in Homo Sapiens and obtained through X-ray crystallography were collected from the PDB as of June 01, 2013. In this set, 15% (16 structures) are bound to GDP, 9.5% (10 structures) are bound to GTP, 63% (66 structures) are bound to GNP, a non-hydrolyzable GTP analog, and 50.5% (53 structures) belong to Ras variants. While there are minor variations on the termini, we focus here on the longest shared stretch of 166 amino acids.

A low-dimensional embedding of the structure space spanned by crystal structures of Ras is obtained through PCA, which is a popular linear dimensionality reduction technique. PCA

finds an orthogonal transformation of points given in some original high-dimensional space that highlights new axes or the PCs, maximizing variance in the projected or transformed data. We conduct PCA here on an ensemble of crystal structures to obtain a low-dimensional embedding of the structure space. The crystal structures are reduced to the C_α atoms or traces prior to the PCA. Visualization and further analysis of this embedding are provided in section 3.

Rather than use all 105 collected structures for PCA, we embed only 46 structures and leave the other ones for further validation. There are two reasons why we do so. First, the selected 46 structures of Ras present the available structural knowledge about Ras in 2009, and these structures were used in a related computational study by Grant and colleagues [7], where PCA was also used to embed the Ras structure space as of 2009. We select here to do the same. Second, the remaining structures present an opportunity to test whether the proposed algorithm explores the structure space in regions not populated by the 46 structures but yet found to be accessible by Ras after 2009. The proposed algorithm, detailed below, is guided by the PCs obtained from PCA on the 46 structures. Therefore, any new structures in regions not spanned by the ones used for the PCA but found to be populated by Ras structures not included in the PCA constitutes novel information generated about the structure space by the algorithm. Of the remaining $105 - 46$ crystal structures, only 41 share a contiguous stretch of 166 amino acids with the rest of the X-ray structures.

2.2 PC-guided Probabilistic Search

The proposed algorithm grows a tree in structure space rooted at a given crystal structure. The tree grows one trajectory at a time. Essentially, a trajectory of consecutive low-energy structures is obtained until it is not possible to do so. When a high-energy structure is encountered, the structure is considered a dead end for the current trajectory, and a new starting structure is selected from any of the ones currently in the tree. The process of obtaining a new path begins anew.

The essential functional core of the algorithm addresses how, given a current structure, to generate the next one in the trajectory. This is implemented by decoupling the geometry from energy. First, given a current structure C, its C_α atoms are first extracted, obtaining the C_α trace T of C. Given T, a perturbation is introduced in order to obtain a new trace $T^{'}$ (we delay details on the perturbation). Once a new trace $T^{'}$ is obtained, a full all-atom conformation is reconstructed by first predicting a backbone conformation and then adding side chains. Backbone atoms are predicted on the newly sampled trace through the BBQ statistical protocol [8]. Our analysis in section 3 shows that the reconstruction is able to recover given experimental structures. Side-chain addition and energetic refinement on a backbone structure is performed through the Rosetta protocol, immobilizing BBQ-reconstructed backbone atoms [13].

The proposed algorithm uses the PCs elucidated by the PCA summarized above to implement the perturbation component and obtain a new trace. The number of dimensions (or PCs) are determined first. We do so by measuring the distortion introduced between traces of the original crystal structures subjected to the PCA and projections of these traces on a given number n of dimensions (PCs are ordered from corresponding highest to lowest eigenvalues, and PCs 1 to n are selected in this order). The analysis detailed in section 3 measures this distortion in terms of RMSD and at three different values of n in order to select a value that minimizes distortion while allowing for a low-dimensional space for exploration.

The selected n PCs present directions of perturbation in the n-dimensional space. The magnitude of perturbation by each PC to the trace of a current structure is determined as follows. The weight w_1 for the first (with highest eigenvalue) PC is sampled uniformly at random in the $[-1, 1]$ range. The weight w_i of any other PC $i > 1$ is then scaled based on the variance associated with that dimension: $w_i = w_1 \cdot \frac{PC_i variance}{PC_1 variance}$. Variance information is available from the eigenvalues returned by the PCA. The purpose of scaling the weights relative to that sampled for the first PC is to maintain the ratio between the ranges spanned by each of the dimensions in the low-dimensional embedding revealed by PCA. Analysis of the ability of this basic perturbation component to generate a trajectory of C_α traces that populates the structure space well is provided in section 3. Analysis of low-energy all-atom structures obtained by the full algorithm is also provided in that section.

3. RESULTS

Two sets of analyses are presented in this paper. We first provide details on the conformational space and associated potential energy surface populated by crystal structures of wildtype and mutant Ras. Using this information, we then provide analysis of the structure space and associated energy surface for Ras as populated by our algorithm.

3.1 Analysis of Space of Crystal Structures

The results of the PCA are shown in Figure 2(a). The left quadrant of the plot in Figure 2(a) shows the accumulation of variance captured by the PCs in order of highest to smallest eigenvalue. This plot shows that linear dimensionality reduction is suitable, capturing 80% of the variance among structural data with 5 dimensions (PCs); even the first three PCs capture about 66% of the variance. This result is in strong agreement with that provided in [7]. The structural space projected on the first two PCs is shown in Figure 2(a). The range of the axes and overall placement of the 46 crystal structures subjected to the PCA are both in good agreement with results reported in [7].

The projections of the 46 crystal structures in Figure 2(a) are color-coded by binding target. Structures bound to GDP are colored in green, whereas those bound to GTP are in red. The remaining 41 crystal structures not used to obtain the PCs are nonetheless projected onto the PC space and colored in blue. PDB ids are shown for as many PDB structures as possible while preserving clarity. The embedded structure space allows drawing a few interesting observations. First, the majority of the GTP-bound structures are located nearby one another, in the bottom right quadrant of the space, whereas the GDP-bound structures span the top left and bottom left quadrants. While some distortion in the embedded space is to be expected, co-localizations of groups of structures are reliable, as the first two PCs capture about 55% of the structural variance. The PC space shows a clear separation among Ras structures based on binding target. Second, the 41 crystal structures not subjected to PCA but projected nonetheless and drawn in blue in Figure 2 are captured well by the existing structure space. As shown, these

structures are largely located in the bottom right quadrant, and their projections reside within the ranges of the PCs obtained by the PCA on the 46 crystal structures.

We additionally examine the clustering of binding site motifs in the 46 crystal structures through the FASST web server [15]. The motif provided to the server consists of residues Q61, G60, A59, A11, G12, G13, V14, G15, K16, S17, T35, A18, V29, F28, N116, A146, S145, D119, K117.

The server conducts essentially a PCA of the binding sites only, and the results of the embedding on the first two PCs are shown in Figure 2(b). Comparison of PDB ids shows that the dots drawn in green in Figure 2(b) correspond to those drawn in green (GDP-bound) in Figure 2(b), suggesting that the binding motif is similar among the GDP-bound structures. The dots drawn in blue and lime in Figure 2(b) correspond to those drawn in red and blue in Figure 2(b), supporting the conclusion that the PCA is capturing well structural differences mainly due to the binding site and co-localizing well in GTP-bound structures. Additionally, the structures with PDB ids 1xcm and 3kkn, color-coded in orange in Figure 2(b), correspond to blue dots with PC1 coordinate around -5 and PC2 coordinate around -3 in Figure 2(a). These structures consistently form their own cluster in both plots. Overall, comparison of the PCA, whether on the entire structure space or only on binding site motifs, allows concluding that PCA captures well differences mainly to binding site and provides a reliable space to use for further analysis and exploration.

3.2 Ras Energy Surface Probed by X-ray Crystallography

In this analysis, we reconstruct the energy surface of wildtype Ras with the information available in the crystal structures. We essentially treat the crystal structures of wildtype and mutant Ras as potentially energetically-stable structures for the wildtype sequence of Ras. As summarized in section 2, we restrict our attention to projections of traces of the crystal structures on n PCs. These projections are then reconstructed in full all-atom detail and energetically refined through Rosetta as detailed in section 2. This approach gives us the opportunity to determine a suitable value for n that allows recovering the original crystal structures and obtain low-energy values for them after refinement.

In Figure 3(a) we show results on the analysis on varying number of PCs. Three settings are considered, using $n \in \{5, 7, 10\}$ PCs. The analysis proceeds as follows. Given n PCs, the projection of a crystal structure on these n PCs is obtained. This projection corresponds to a deformed C_α trace of the crystal structure, as structural information available in the rest of the PCs (after n) is discarded. The RMSD between this deformed trace and the full trace is calculated, in order to measure the distortion introduced by cutting down the number of PCs to n. The histogram of RMSDs over all crystal structures is shown in Figure 3(a). Three different colors are used for the three settings of n.

As expected, a smaller number of PCs reduces the dimensionality of the structure space but also introduces more distortions (higher RMSDs). The distributions obtained when using only 5 or only 7 PCs are bimodal. This is due to the fact that structures corresponding to low RMSDs in the distribution are those 46 used to obtain the PC space. The higher RMSDs correspond to the additional 41 structures not subjected to PCA. If 10 dimensions are used, then the bimodality disappears, and on average a low RMSD distortion is present. It is worth noting that larger RMSDs are obtained for 5 structures in the set not subjected to PCA. This is due to a structural difference in a random coil not present in any other Ras structure. Consequently, the PC space is not able to encode this structural difference and any number of PCs would not suffice to rebuild any of these 5 structures. Taken together, this analysis justifies employing $n = 10$ PCs in the rest of our analysis and the results related for the exploration by the proposed search algorithm.

Each crystal structure, after being threaded onto the wildtype sequence, is reduced to its (distorted) projection on $n = 10$ PCs. The distorted trace is rebuilt into an all-atom structure and refined (detailed in section 2). A potential energy value is thus associated with each structure. The RMSD between the trace of each original crystal structure and the trace of the rebuilt and refined version on the wildtype sequence is measured, and the histogram of such RMSDs in Figure 3(b) shows that the RMSDs are overall small, less than 1Å. Figure 3(b) illustrates this by superimposing a rebuilt over an original structure.

Figure 4 replots the PC-projected structure space, color-coding each structure according to its all-atom (Rosetta) potential energy value after the rebuilding and energy refinement process. Figure 4 shows a lot of low-energy structures and some higher-energy ones, suggesting a richness of thermodynamically-accessible structures to wildtype Ras.

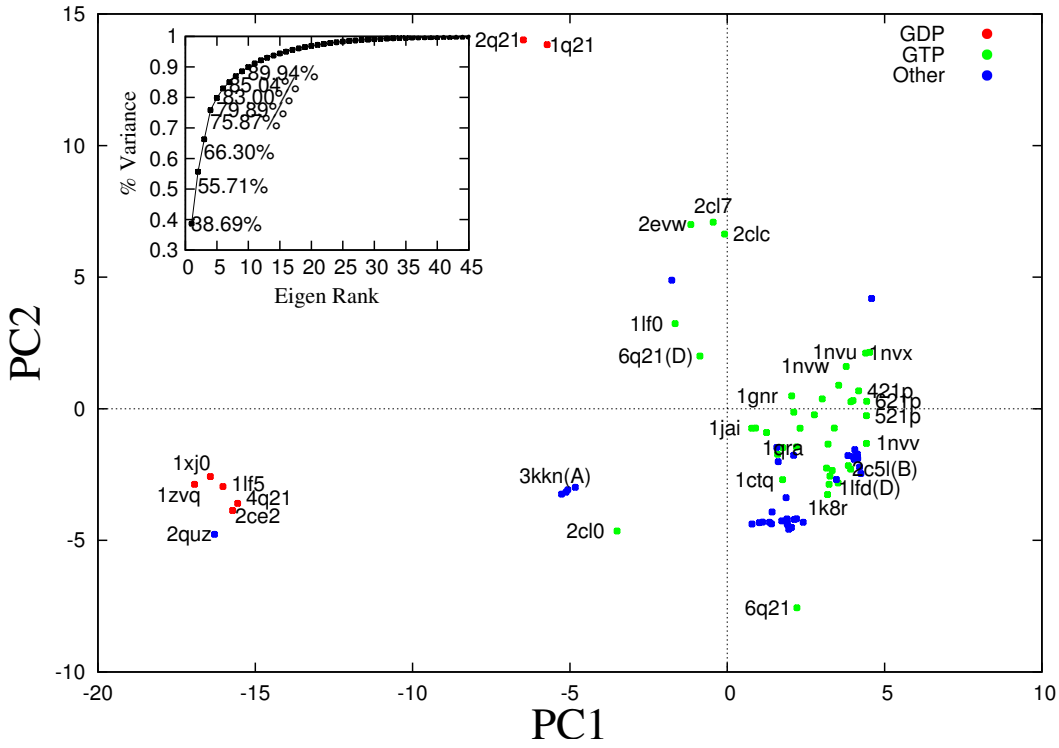
3.3 PCA-guided Exploration of Wildtype Ras Structure Space

The promise of the geometrical exploration in the proposed algorithm in further populating the structure space of wildtype Ras, as guided by the top $n = 10$ PCs, is illustrated in Figure 4. An initial structure is selected (PDB id 1agp here), and the process of sampling weights for each of the n PCs and using them to generate a new C_α trace from a current one is repeated to obtain a trajectory of consecutive traces. This trajectory of 1000 traces, shown in Figure 4, makes the case that the basic component of guiding sampling of the structure space by the PCs can potentially further populate the space and reveal new structures not currently probed by experimental techniques.

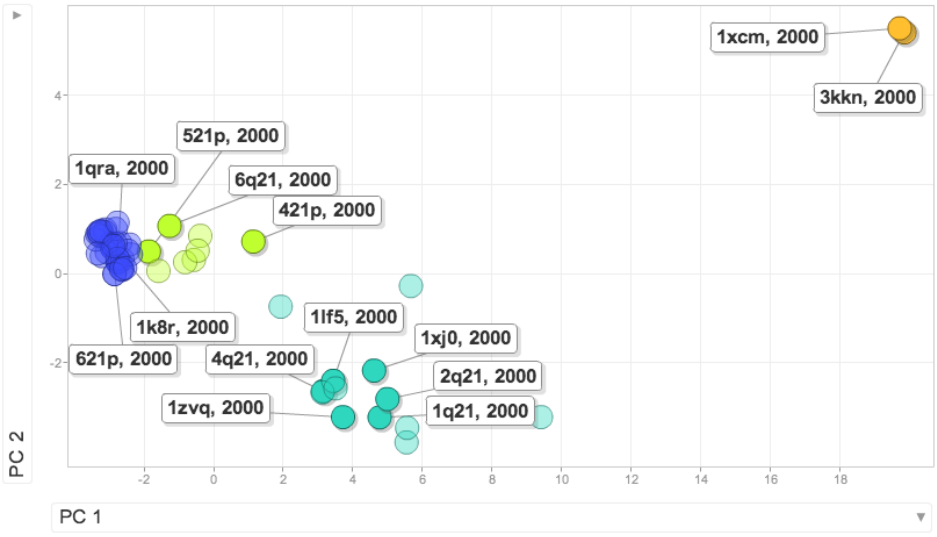
The shown trajectory provides geometrical information about the space. Sampled traces may be energetically unfavorable, and the rest of the components of the algorithm are needed to sample low-energy regions of the wildtype Ras structure space. We note that though the explored space is visually drawn on the first two PCs, all $n = 10$ PCs are used, resulting in a 10-dimensional space being explored.

The full algorithm is now used to explore low-energy regions of the structure space. As detailed in section 2, a (rebuilt and refined) crystal structure is used as a starting point. A trajectory of consecutive structures is generated until a high-energy (higher than 100 kcal/mol of start structure) one is encountered, at which point exploration begins anew from any structure in the tree. Figure 5 shows the results of two runs, one initiated from a GTP-bound reconstructed structure (PDB id 1qra) and another from a GDP-bound reconstructed structure (PDB id 4q21). The 1000 structures obtained by each run are projected onto the top two PCs and superimposed over the crystal structures. Structures are color-coded by potential energy values.

Comparison of the two settings shows that the initial start-

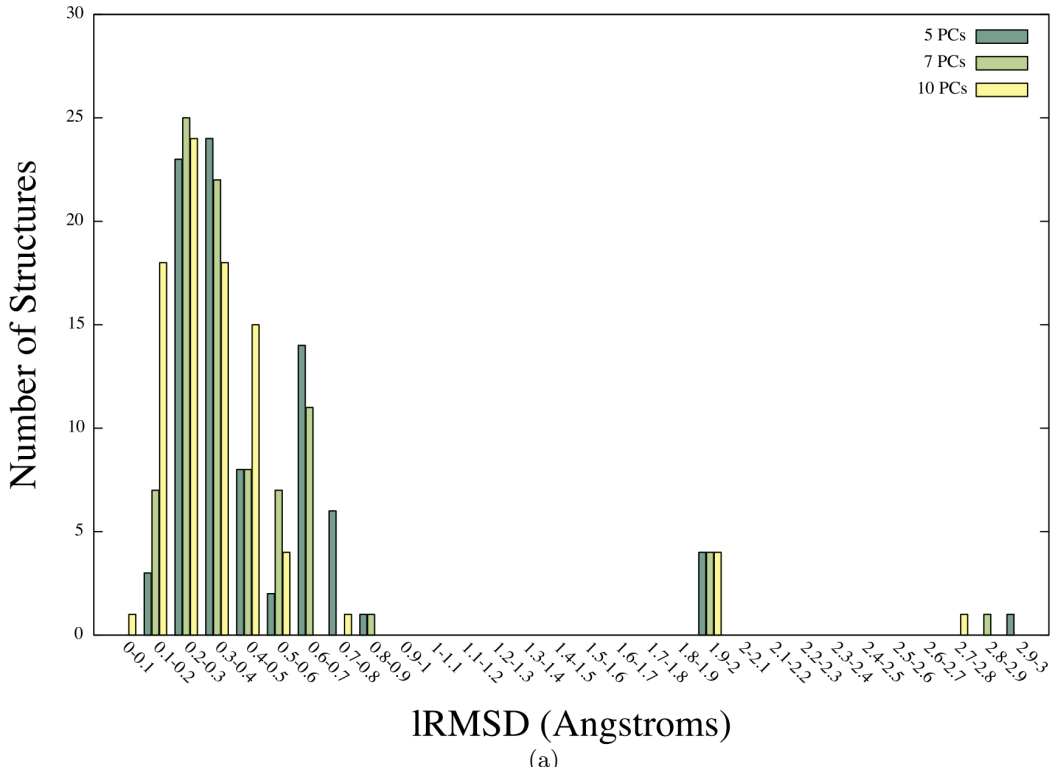


(a)

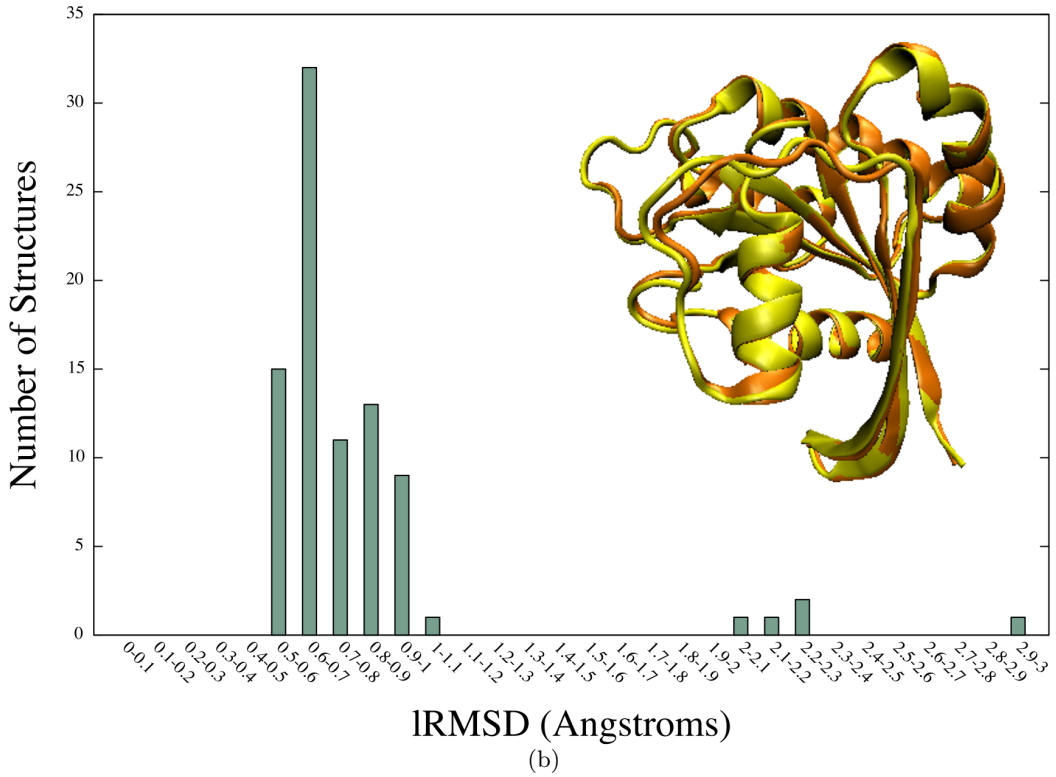


(b)

Figure 2: (a) The structure space spanned by the 46 crystal structures of Ras available in the PDB in 2009 is shown projected on the first two PCs. The accumulation of variance is shown on the top left quadrant. These structures are color-coded based on binding target, green for GTP-bound and red for GDP-bound. The additional crystal structures added to the PDB since 2009 are projected onto the same PC space and are color-coded in blue. (b) PCA of the binding site motifs in Ras crystal structures is provided from the FASST server [15].



(a)



(b)

Figure 3: (a) Histogram shows distribution of RMSDs between the trace of each crystal structure and its projection on n PCs. Settings $n \in \{5, 7, 10\}$ are shown in different colors. (b) Histogram shows distribution of RMSDs between traces of original crystal structures and traces of rebuilt structures. The embedded illustration obtained through VMD superimposes the original crystal structure under PDB id 1agp, drawn in orange, over the rebuilt structure, drawn in yellow.

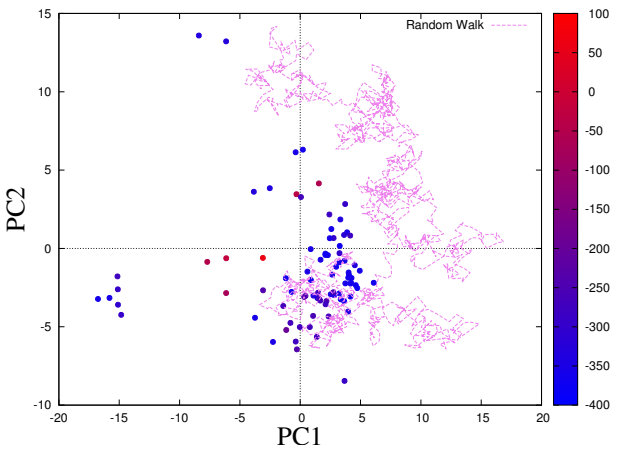


Figure 4: The dots represent rebuilt and refined crystal structures threaded onto the wildtype sequence, color-coded by their all-atom energies. Consecutive traces obtained in a random walk are superimposed over the refined crystal structures.

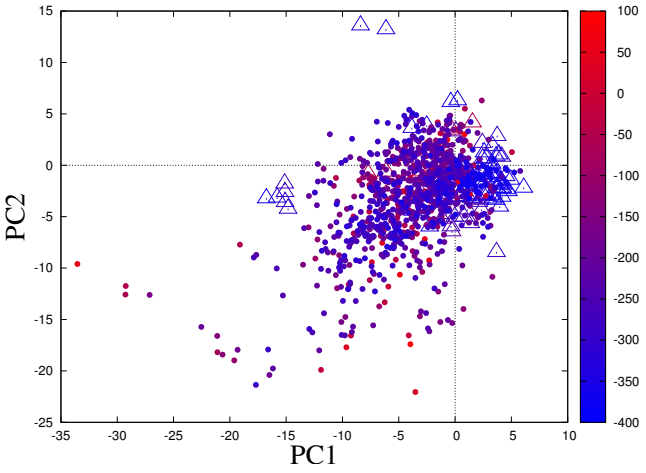
ing point does not significantly affect the exploration capability. The bottom left quadrant is slightly better populated when initiated from the GDP-bound structure, which itself resides in this quadrant. Low energies are obtained in each setting, however, suggesting that the energy surface of wildtype Ras is flat and contains many stable structures. While the top left and bottom right corners of the space are not well populated in each setting, in agreement with the lack of experimental structures in these regions, as well, further sampling may be needed to determine whether these regions are indeed unfeasible for the wildtype sequence.

We quantify the ability of the algorithm to cover the structure space of Ras captured by the crystal structures in Table 3.3. We do so as follows. If the closest structure sampled by the algorithm to a given crystal structure is below a given RMSD threshold, the crystal structure is considered captured. We consider three different RMSD thresholds of 1.0, 2.0, and 2.5Åand report the percentage (%) of crystal structures captured according to this definition. We show separate statistics on the X-ray structures used for the PCA and those withheld for validation. Three different settings are considered, where the structures sampled in each of the two runs (from PDB id 1qra and 41q21) are kept separate and then combined. Taken together, the above analysis and that shown in Table 3.3 make the case that novel low-energy structures are obtained in each setting, suggesting the structure space of Ras is rich and holds potentially valuable information for further analysis.

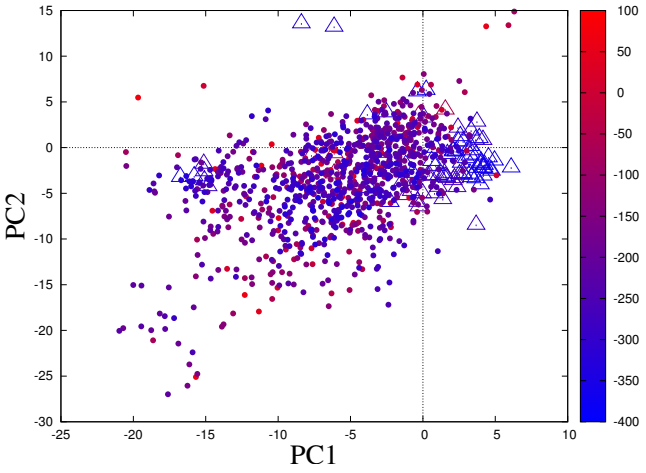
4. CONCLUSION

The approach described in this paper to associate a low-dimensional structure space and potential energy surface with available crystal structures of Ras suggests that the wildtype Ras structure space is flat and particularly ripe for conformational selection by few point mutations with no deleterious effects for folding. The PC-guided search algorithm proposed here to further populate the wildtype Ras structure space is promising, revealing novel structures and accessing new low-energy regions of the structure space.

This work is a first step towards a comprehensive map of



From PDB id 1qra (GTP-bound)



From PDB id 4q21 (GDP-bound)

Figure 5: 1000 low-energy structures obtained when starting from the reconstructed structure under PDB id 1qra (top pannel) or PDB id 4q21 (bottom pannel) are drawn as circles color-coded by potential energy. These are superimposed over the crystal structures drawn as triangles.

the sequence-structure space of Ras. While the work here focuses on wildtype Ras, the investigation can be repeated on other selected variants to show, for instance, how the landscape changes upon mutations. This work opens the way for further studies focused on extracting information from computed structures on novel binding sites, mechanisms, and even interaction partners. Such information can be valuable to improve our understanding of how Ras mutants and their interacting partners affect pathways contributing to cancer.

Acknowledgment

This work is supported in part by NSF CCF No. 1016995 and NSF IIS CAREER Award No. 1144106. We thank B. Olson for his help with scripting and data gathering and B. Grant for sharing with us the 46 Ras crystal structures.

5. REFERENCES

[1] I. Al-Bluwi, M. Vaisset, and T. Simeon. Coarse-grained elastic networks, normal mode analysis

Table 1: The percentage of X-ray structures captured under each setting is measured here, using three different RMSD thresholds. RMSD is measured over C_α atoms.

	X-ray structures used for PCA			X-ray structures withheld from PCA		
	1.0Å	2.0Å	2.5Å	1.0Å	2.0Å	2.5Å
4q21	91.30%	100.00%	100.00%	87.50%	97.50%	97.50%
1qra	95.65%	100.00%	100.00%	85.00%	97.50%	97.50%
Combined	95.65%	100.00%	100.00%	87.50%	97.50%	97.50%

and robotics-inspired methods for modeling protein conformational transitions. In *Comput Struct Biol Workshop (CSBW)*, pages 40–47, Philadelphia, PA, October 2012. IEEE.

[2] H. M. Berman, K. Henrick, and H. Nakamura. Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.*, 10(12):980–980, 2003.

[3] J. F. Díaz, B. Wroblowski, and Y. Engelborghs. Molecular dynamics simulation of the solution structures of Ha-ras-p21 GDP and GTP complexes: flexibility, possible hinges, and levers of the conformational transition. *Biochemistry*, 34(37):12038–12047, 1995.

[4] J. F. Díaz, B. Wroblowski, J. Schlitter, and Y. Engelborghs. Calculation of pathways for the conformational transition between the GTP- and GDP-bound states of the Ha-ras-p21 protein: calculations with explicit solvent simulations and comparison with calculations in vacuum. *Proteins*, 28(3):434–451, 1997.

[5] A. Fernández-Medarde and E. Santos. Ras in cancer and developmental diseases. *Genes Cancer*, 2(3):344–358, 2011.

[6] C. K. Foley, L. G. Pedersen, P. S. Charifson, T. A. Darden, A. Wittinghofer, E. F. Pai, and M. W. Anderson. Simulation of the solution structure of the H-ras p21-GTP complex. *Biochemistry*, 31(21):4951–4959, 1992.

[7] B. J. Grant, A. G. Alemayehu, and J. A. McCammon. Ras conformational switching: Simulating nucleotide-dependent conformational transitions with accelerated molecular dynamics. *PLoS Comp Biol*, 5(3):e1000325, 2009.

[8] D. Gront, S. Kmiecik, and A. Kolinski. Backbone building from quadrilaterals: a fast and accurate algorithm for protein backbone reconstruction from alpha carbon coordinates. *J. Comput. Chem.*, 28(29):1593–1597, 2007.

[9] N. Haspel, M. Moll, M. L. Baker, W. Chiu, and K. L. E. Tracing conformational changes in proteins. *BMC Struct. Biol.*, 10(Suppl1):S1, 2010.

[10] W. Humphrey, A. Dalke, and K. Schulten. VMD - Visual Molecular Dynamics. *J. Mol. Graph. Model.*, 14(1):33–38, 1996. http://www.ks.uiuc.edu/Research/vmd/.

[11] L. Jaillet, F. J. Corcho, J.-J. Perez, and J. Cortes. Randomized tree construction algorithm to explore energy landscapes. *J. Comput. Chem.*, 32(16):3464–3474, 2011.

[12] A. E. Karnoub and R. A. Weinberg. Ras oncogenes: split personalities. *Nat Rev Mol Cell Biol*, 9(7):517–531, 2008.

[13] K. W. Kaufmann, G. H. Lemmon, S. L. DeLuca, J. H. Sheehan, and J. Meiler. Practically useful: What the rosetta protein modeling suite can do for you. *Biochemistry*, 49(14):2987–2998, 2010.

[14] S. Kirillova, J. Cortes, A. Stefaniu, and T. Simeon. An nma-guided path planning approach for computing large-amplitude conformational changes in proteins. *Proteins: Struct. Funct. Bioinf.*, 70(1):131–143, 2008.

[15] M. Moll, D. H. Bryant, and L. E. Kavraki. The LabelHash algorithm for substructure matching. *BMC Bioinf*, 11(1):555, 2010.

[16] K. Molloy and A. Shehu. A robotics-inspired method to sample conformational paths connecting known functionally-relevant structures in protein systems. In *Comput Struct Biol Workshop (CSBW)*, pages 56–63, Philadelphia, PA, October 2012. IEEE.

[17] S. Nath, S. Thomas, E. Chinwe, and N. M. Amato. A multi-directional rapidly exploring random graph (mrrg) for protein folding. In *ACM Conf on Bioinf and Comp Biol (BCB)*, pages 44–51, Orlando, FL, October 2012.

[18] B. Olson, K. Molloy, and A. Shehu. Enhancing sampling of the conformational space near the protein native state. In *BIONETICS: Intl. Conf. on Bio-inspired Models of Network, Information, and Computing Systems*, Boston, MA, December 2010.

[19] B. S. Olson, K. Molloy, S.-F. Hendi, and A. Shehu. Guiding search in the protein conformational space with structural profiles. *J. Bioinf. and Comput. Biol.*, 10(3):1242005, 2012.

[20] A. Shehu. An ab-initio tree-based exploration to enhance sampling of low-energy protein conformations. In *Robot: Sci. and Sys.*, pages 241–248, Seattle, WA, USA, 2009.

[21] A. Shehu. Conformational search for the protein native state. In H. Rangwala and G. Karypis, editors, *Protein Structure Prediction: Method and Algorithms*, chapter 21. Wiley Book Series on Bioinformatics, Fairfax, VA, 2010.

[22] A. Shehu. Probabilistic search and optimization for protein energy landscapes. In S. Aluru and A. Singh, editors, *Handbook of Computational Molecular Biology*. Chapman & Hall/CRC Computer & Information Science Series, 2013.

[23] 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.

[24] S. Thomas, G. Song, and N. M. Amato. Proten folding by motion planning. *J. Phys. Biol.*, 2:S148–S155, 2005.