Coarse-graining of Protein-Nanoparticle Interactions

David Power, Stefano Poggio, Hender Lopez and Vladimir Lobaskin

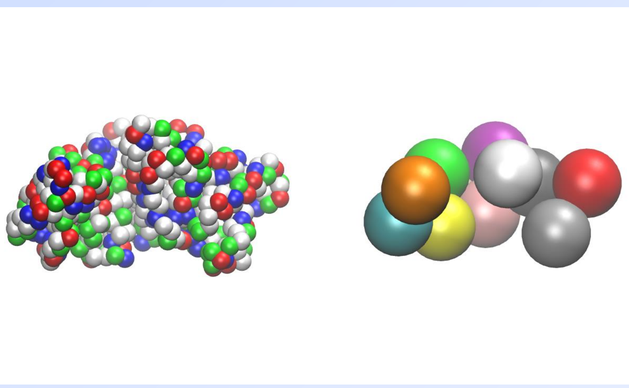
School of Physics, University College Dublin, Belfield, Dublin 4, Ireland

Introduction

Even with the exponential growth in modern computing power, it is still computationally infeasible to simulate a fully atomistic environment with more than a few proteins for more than a few nano-seconds. One way to overcome this problem is to come up with simpler representations of the components of the simulation. To this end, we aimed to reduce the representation a the protein to as few beads as possible while still maintaining the overall shape and adsorption characteristics of the original protein.

Methods

We used VMD Coarse Grain Builder (CGB) to turn a fully atomistic representation of the protein into a model with a much reduced number of representative beads, that still preserves the overall shape of the protein. For this we used CGBs shape-based method, where a neural network learning algorithm is used to determine the placement of neurons (or CG beads). The CG beads have masses correlated to the clusters of atoms which the beads are representing [1]. If the PDB information for the protein structure was unavailable, we used the I-TASSER suite of protein stucture and function prediction tools to model some possible confirmations of the protein [2].



Transformation of Human Serum Albumin (1n5u) from a fully atomistic model to an 11 bead model.

Methods (Continued)

Once we had this simplified representation of the protein, we applied a pseudo-random potential to each bead. We wanted these potentials to give the protein the same adsorption energy characteristics (i.e. adsorption energy per angle heat-map) that a more complicated model would give. For this we used a genetic algorithm to minimize the difference between the heat-maps generated by our population of bread/mutated proteins and a reference heat-map [3]. This involved minimizing the fitness function:

$$k = \sum_{\theta,\phi} (U(\theta,\phi) - U(\theta,\phi)_{\mathrm{ref}})^2$$

Where the generated potential energy is given by:

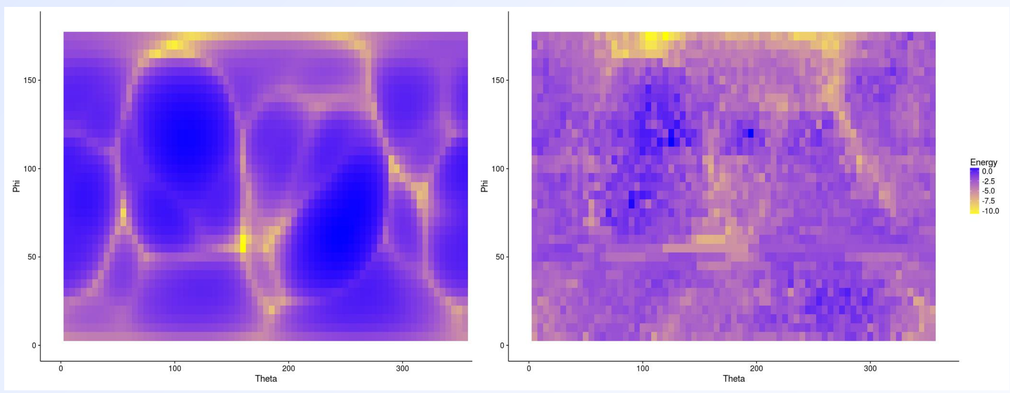
$$U(\phi,\theta) = -kT \sum^{n} ln\left(\frac{3}{(R_n+s)^3 - R_n^3}\int_{R_n}^{R_n+s} r^2\, e^{-U_n(r_n)/kT}\right)$$

During both the breeding and mutation phases of the algorithm, mathematical functions were applied that would ensure that the potentials remained continuous and would approach zero away from the source.

Results

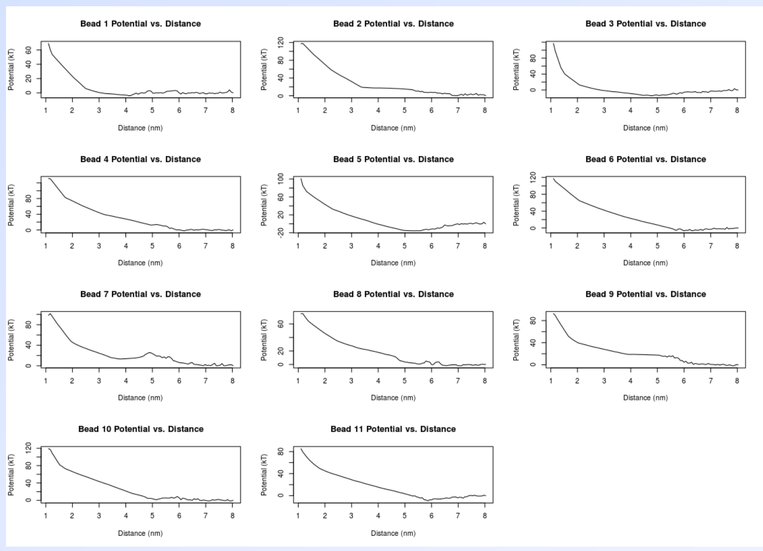
Below is the result of running the GA for approximately 50,000 generations (or about 5 days on a 10-core, 2.4-GHz machine). The fitness function value plateaued after around 10,000 generations.

Reference heat-map (left) for HSA compared to new GA minimized heat-map (right). The main valleys/potential wells of the energy landscape are still captured in the simplified model.



Results (Continued)

Below we have the 11 unique potentials generated for the 11 beads of the simplified HSA model. They are all continuous and approach zero away from the origin.



Conclusion

This method seems to be a feasible way to reduce the constituent components of a protein which would help with computational simulations. One of the drawbacks of this method is that the proteins are now rigid, which is unlikely to match real world adsorption characteristics.

Next Steps

Would like to apply a Boltzman weighting to the fitness function so that the valleys for the energy map (e.g. the important parts we are interested in) have a larger impact on fitness selection. We would also like to reduce the number of pixels in the generated heat-map. We currently keep the same number of pixels as in the reference heat-map. Reducing the number of pixels would increase the speed of computation while, we believe, only reduce the accuracy from it's current level by a small amount.

References

1 Shih, A. Y., Freddolino, P. L., Arkhipov, A., & Schulten, K., J. Struct. Biol, 2007, 157(3), 579–592, 10.1016/j.jsb.2006.08.006

2 Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y., Nat. Methods, 2014, 12(1), 10.1038/nmeth.3213

3 Lopez, H., & Lobaskin, V., J. Chem. Phys, 2015, 143(24), 10.1063/1.4936908

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 686098

European Commission | Horizon 2020 European Union funding for Research & Innovation