Proceedings of the ***Conference of the Hellenic Society for Computational Biology and Bioinformatics (HSCBB21)***
Hellenic Pasteur Institute, Athens, Greece, 10-11 December 2021

Geometric modeling of coiled-coils using Isambard : the case of the RM6 variant of the Repressor of Primer protein.

Magdalini Chatzopoulou, Nicholas M. Glykos
Department of Molecular Biology and Genetics, Democritus University of Thrace, University Campus, Alexandroupolis, 68100, Greece
∗Correspondence to: magdalini.chatzopolou@gmai.com , glykos@mbg.duth.gr

***Abstract*— Retro-proteins are molecules with reversed amino acid sequences compared to their parents. The inversion of a protein sequence produces a new polypeptide chain that does not exhibit any homology with its parent and thus its foldability is unknown. Results from studies of such retro-proteins have been contradictory but recent experiments on the retro-RM6 protein (a deletion variant of the ROP protein) suggested that rRM6 is stable and with a similar fold to its parent. In order to investigate retro-proteins computationally, ISAMBARD represents a potential tool for geometrical modeling, model evaluation and parameter optimization for coiled coils. In this preliminary report, we examine the ability of geometrical modeling to determine the structure of the RM6 protein (PDB ID: 1QX8) and show that we can obtain refined models exhibiting RMSD values (from the crystallographic structure) of only ~1.15Å for 196 residues of the bundle. These results suggested that geometrical modeling via the ISAMBARD tool could be employed to generate potentially useful models for the retro-isomer of RM6.**

I. Introduction

ROP protein (PDB ID: 1ROP) constitutes a well studied representative of an anti-parallel four-α-helix bundle with each monomer consisting of two anti-parallel α-helices connected by a short loop. One of its variants, RM6, is a stable and regular homotetrameric left-anti-parallel helix bundle with a five residue deletion (DADEQ) in the turn region which restores the heptad motif. Kefala et. al (2021) used circular dichroism spectroscopy, size exclusion chromatography combined with multi-angle laser light scattering, and small angle X-ray scattering to investigate the structural properties of the rROP and rRM6 retro-isomers [1]. Their studies suggested that rRM6 exhibited high similarity to its parent on a secondary structure and oligomerization level, in addition to being slightly less compact compared to RM6. On the other hand, the rROP protein displayed an unclear oligomerization state and a disordered, molten globule state. Molecular replacement attempts on crystallographic data obtained from the rRM6 crystals [2] failed to allow a complete structure determination, implying potential differences between the retro-isomer and its parent. In order to obtained possibly useful models for molecular replacement calculations, geometrical modeling could be employed that utilizes our prior knowledge on coiled coil geometry in order to *de novo* build protein backbones. This method does not require an available resolved structure or sequence homology which proves useful when modeling a retro-protein. The tools CCBuilder and ISAMBARD [3,4] offer the parameters and approaches necessary to solve the optimization problem of coiled coil parameterization. More specifically, ISAMBARD offers geometrical parameters that describe coiled coil features i.e. superhelical radius, pitch, φCα angles, helix orientation, superhelix handedness, z-shift and metaheuristic methods to search the parameter space.

II. Methods

ISAMBARD requires Coiled Coil specifications, parameter ranges and amino acid sequences. The value ranges of the parameters for optimization were obtained from previous research on tetrameric coiled coils [4]. In general, the metaheuristic methods employed involve a population of candidate solutions which are altered and then assessed based on an evaluation function in an iterative manner. The models with the best score are used to initiate the next generation until a specified number of cycles has been met. In this case, BUFF (Bristol University Docking Engine Force Field), which is an empirical free-energy force field is integrated to ISAMBARD to evaluate the generated models [4]. The sequences used for anti-parallel models were selected based on the heptad motif of the RM6 protein. For the all-parallel structures, models with all possible sequence-parameter combinations were generated and the one with the lowest energy was selected. For model evaluation, the RMSD values for all the generated models were calculated before and after refinement with the GalaxyRefineComplex web server [5].

III. Results

Figure 1 shows the general form of the results obtained from ISAMBARD : the horizontal axis corresponds to the various possible heptad registers (*a*-*g*) of the residues, the vertical axis is the computed energy. Note that for each possible register the program also refines a multitude of other parameters such as exact helix orientation, z-shift, etc. For the case illustrated in Figure 1, we compare the energies obtained from the antiparallel arrangement of RM6 (known to be the correct one, black curve), with a hypothetical parallel arrangement (red curve) of the helices. The results clearly indicate the presence of one and only one pronounced energy minimum for the *d* register which is only present when the (correct) antiparallel helical arrangement is used.

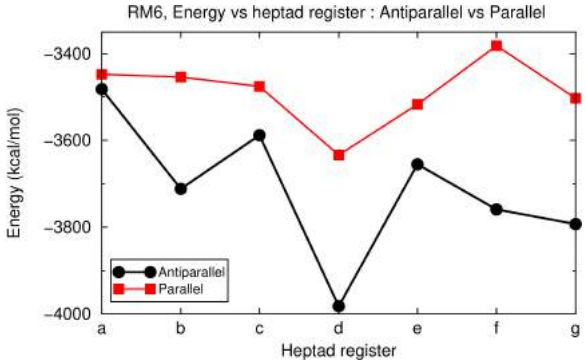

Figure 1 : Comparison of the energies obtained from the antiparallel vs parallel arrangements of RM6.

The ability of geometric modeling to correctly differentiate between different solutions of the helix packing in bundles is better illustrated in Figure 2. In this figure we compare two 'normal' (antiparallel) arrangements with the only difference between them being the hand of the superhelical twist of the four helices, with the left hand known from the experiment to be the one observed. Again, geometric modeling gives a single, clear, and convincing solution corresponding to the correct *d*-antiparallel-left-handed solution.

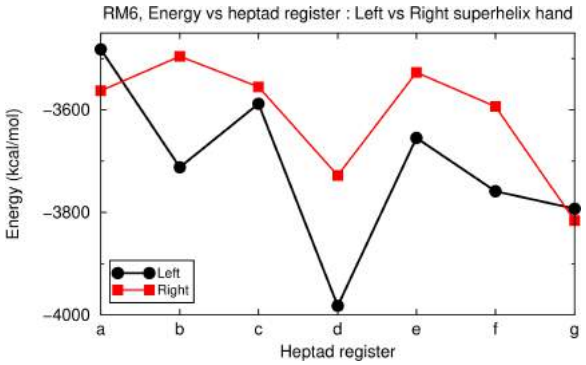

Figure 2 : Comparison of the energies obtained for a left-handed vs right-handed antiparallel arrangement of RM6.

The ability of this protocol to locate geometrically acceptable solutions is not, however, adequate. Accuracy at the atomic level is needed if these models are to be used as initial models in other calculations. To examine this issue, we calculated the RMSD between the experimental structure and the structure obtained from the *d*-antiparallel-left-handed solution discussed above. The initial RMSD over 192 Cα atoms was only 1.8Å, which dropped to an impressive RMSD of only 1.15Å after energy minimization with Galaxy[5]. Figure 3 shows a superposition between the experimental and the modeled structure of RM6. Clearly, this is an impressive agreement, especially when considering that it is the result of a geometric and not knowledge-based modeling. Having said that, it must be noted that geometric modeling is powerful enough to be able to find convincing solutions even when initiated from completely wrong initial parameters (data not shown). This rightfully necessitates the application of a fair amount of meticulousness and alertness from its users.

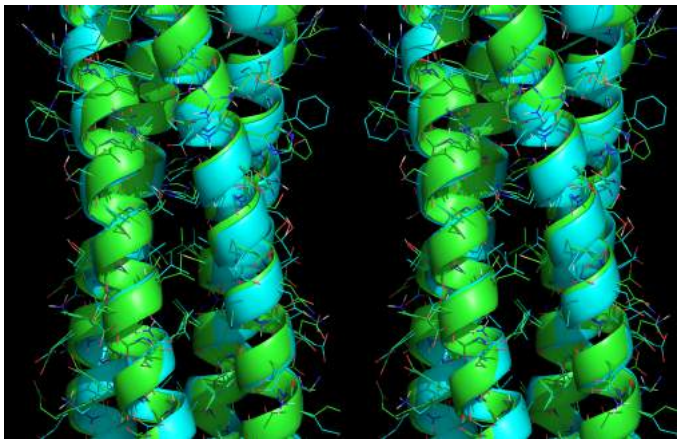

Figure 3 : Wall-eyed stereo view of the superposition between the experimental (blue) and modeled (green) RM6 with side chains shown using a liquorice representation. Only a portion of the structure is shown for clarity.

IV. Conclusions & Future work

We have shown that for the case of native RM6 geometric modeling using ISAMBARD can produce models of quality and accuracy sufficient even for demanding calculations such as crystallographic molecular replacement. Application of the procedure to an unknown structure such as rRM6 is not, however, straightforward. To start with, the target protein may not even form a canonical bundle which if true will invalidate the procedure right from the start. Additionally, geometric modeling is sufficiently powerful to be able to produce structures with reasonable packing, but with an offset register of the helices, which further complicates the identification of a putative correct solution. Given these limitations, one possible approach would be to produce and test a large number of rRM6 models for their ability to allow crystallographic structure determination to proceed to completion.

References

[1] Kefala A, Amprazi M, Mylonas E, Kotsifaki D, Providaki M, Pozidis C, et al. Probing Protein Folding with Sequence-Reversed α-Helical Bundles. Int J Mol Sci. 2021 Jan;22(4):1955.

[2] Kefala A, Kotsifaki D, Providaki M, Amprazi M, Kokkinidis M. Expression, purification and crystallization of a protein resulting from the inversion of the amino-acid sequence of a helical bundle. Acta Crystallogr Sect F Struct Biol Commun. 2017 Jan 1;73(Pt 1):51–3.

[3] Wood CW, Heal JW, Thomson AR, Bartlett GJ, Ibarra AÁ, Brady RL, et al. ISAMBARD: an open-source computational environment for biomolecular analysis, modelling and design. Bioinforma Oxf Engl. 2017 Oct 1;33(19):3043–50.

[4] Wood CW, Bruning M, Ibarra AÁ, Bartlett GJ, Thomson AR, Sessions RB, et al. CCBuilder: an interactive web-based tool for building, designing and assessing coiled-coil protein assemblies. Bioinforma Oxf Engl. 2014 Nov 1;30(21):3029–35.

[5] Heo, L., Lee, H. & Seok, C. GalaxyRefineComplex: Refinement of protein-protein complex model structures driven by interface repacking. Sci Rep 2016, 6, 32153.