Constrained Modelling of an Intermediate Filament Dimer

Dmytro Guzenko, Sergei V. Strelkov

Laboratory for Biocrystallography, Department of Pharmaceutical and Pharmacological Sciences, KU Leuven, Belgium, Dmytro.Guzenko@pharm.kuleuven.be

Introduction

Intermediate filaments (IF) are the principal components of the cytoskeleton in most animal cells. They perform essential mechanical functions within the cell, and mutations in IF genes are responsible for currently incurable diseases, including muscle, heart, skin and neurological disorders [1]. Hence a better understanding of the IF structure and function is an ongoing demand, and computational tools are required to aid the experimental process.

The elementary building block of IFs is an elongated dimer. A change in solution environment such as its ionic strength can lead to self-assembly of dimers into larger oligomers and ultimately to long, about 10nm -wide filaments. A high quality dimer structure is therefore a prerequisite to successful *in silico* modelling of the IF assembly process. Crystallographic data for a number of fragments of the IF protein vimentin is available [2] and can be used to model an elementary IF dimer, with computer simulations to fill-in the blanks. EM and cross-linking experiments provide additional constraints on the IF architecture – although far from exhaustive, they can be employed to guide the computational methods as well. The constrained *ab initio* modelling of the IF proteins utilises their highly regular structure and greatly reduces the conformational search space in further applications.

Results

We have implemented the 'divide-and-conquer' protein folding approach by customizing various Rosetta suite protocols [3], validating on a vimentin full-length dimer. We employ individual tactic for each domain.

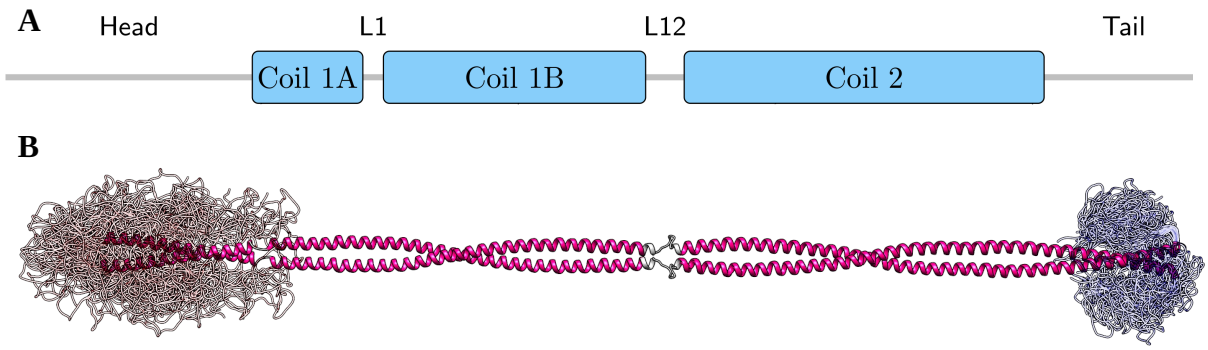

Figure 1: (A) Domain diagram and (B) full atomic model of human vimentin dimer.

1. For many coiled-coil (CC) proteins, including IF fragments, experimental atomic resolution data are available. Correspondingly, for the central CC domain (the 'rod') of the IF dimer we use constrained folding, based on the symmetric Fold-And-Dock protocol. The soft distance constraints are imposed to guide the Monte Carlo search towards the regular CC structure, with the Rosetta energy function determining the details of the conformation.

2. With no experimental constraints for the linker L1 and L12, we used purely computational approach, selecting a representative model from the best-energy cluster of the Rosetta Fold-and-Dock output.

3. The intrinsically disordered head and tail domains were built using a customized 'Floppy Tail' protocol of the Rosetta suite, producing a 'cloud' of conformations with probabilistic interpretation.

Methods

Distance constraints. The guiding distance constraints are automatically predicted based on the sequence profile using a Least Squares Support Vector Machine (LS-SVM) [4] trained on all coiled coil fragments available in Protein Data Base (PDB). We were able to achieve a reliable prediction of Ca-Ca distances between the opposing residues in a dimer with a standard deviation of 0.9A.

Divide-and-conquer folding. Considering the 'linear' nature of the rod domain with no long-range interactions in a dimer, we split the modelling problem into short overlapping segments and perform the folding independently.

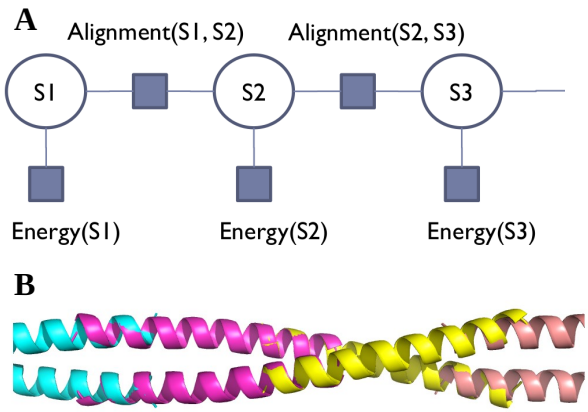

Figure 2: (A) Factor graph of the Markov model and (B) sample result of the divide-and-conquer protein folding

To determine the optimal segmentation points we have implemented a structural diversity indicator: for all sets of sequence fragments that define conformational search space in the folding algorithm, their structural diversity is estimated by calculating Renyi

entropy of the dihedral angles distribution. Consequently, the segmentation points are selected in the regions with lower than average diversity. This approach ensures that termini of a modelled segment are 'stable', facilitating accurate merging of the fragments obtained by independent simulations.

The complete rod domain is reassembled via a probabilistic Markov Model, which derives an 'optimal path', employing Rosetta energy scores and quality of structural alignment of the overlapping segments.

The full dimer model was manually validated using available crystal structures of the vimentin fragments. The framework designed allows us to automatically produce the dimers of other IF classes, which all share the same tripartite structure as vimentin.

This research was supported by the KU Leuven (grant OT13/097 to SVS), by Research Foundation Flanders FWO (grant G070912N to SVS) and by the EU COST Action 'Nanonet'.

References

1. M. B. Omary (2009) "IF-pathies": a broad spectrum of intermediate filament-associated diseases, *J. Clin. Invest.*, **119(7)**:1756–1762.

2. A.A. Chernyatina et al. (2015) Intermediate Filament Structure: The Bottom-up Approach, *Curr. Opin. Struct. Biol.*, **32**:65-72.

3. K.W. Kaufmann et al. (2010) Practically Useful: What the Rosetta Protein Modeling Suite Can Do for You, *Biochemistry* **49(14)**:2987-2998.

4. J.A.K. Suykens, J. Vandewalle (1999) Least squares support vector machine classifiers, *Neural Process. Lett.*, **9(3)**:293–300.