Assessing the Stability of Protein Complexes within Large Assemblies

Frédéric Cazals, Tom Dreyfus

Assessing the Stability of Protein Complexes within Large Assemblies *

F. Cazals †and T. Dreyfus ‡

June 19, 2009

Abstract

This work introduces (i) toleranced collections of balls to represent protein assemblies known with uncertainties (ii) a method to highlight stable complexes within such assemblies. This method is used to check the coherence between Tandem Affinity Purification data and plausible reconstructions of the Nuclear Pore Complex.

1 Large assemblies and tandem affinity purification data: complexes or mixtures?

Structural genomics projects, in particular those exploiting Tandem Affinity Purification (TAP), have revealed remarkable features of full proteomes [G06]. While these insights are essentially of combinatorial nature, that is a number of proteins are known to interact within a complex, leveraging this information will require building three dimensional models of these assemblies. Such an endeavour has recently been completed for the Nuclear Pore Complex, for which plausible reconstructions have been computed from different experimental data, including TAP data [A07a,A07b]. The reconstruction procedure used is reminiscent from NMR in that restraints encoding the coherence w.r.t. experimental data have been used to defined the optimization criterion. Yet, a full synergy between TAP data and the 3D reconstruction is not at play for two reasons. First, the models built are qualitative, that is no resolution whatsoever is claimed. Second, the connexion between the reconstruction and the TAP data has not been elucidated. In particular, deciding whether proteins seen on a gel of a TAP experiment correspond to a single complex or a mixture of complexes within the assembly is not addressed. This work provides methods answering both questions.

2 Tracking stable complexes in uncertain assemblies

Consider a large protein assembly. As advocated in [A08], because of the uncertainties attached to the experimental data used to reconstruct such an assembly, the resulting models are likely to be uncertain. To account for these uncertainties, we propose to use toleranced models, such a model being a collection of balls whose radii are not fixed but instead span certain ranges. More precisely, a toleranced ball is defined by a triple (c, rmin, rmax), with c the center, rmin and rmax the extreme values for its radius, with rmax¿=rmin. That is, given a real number l in [0,1], we may interpolate between these values to define r(l) = rmin + l (rmax-rmin). The ball of radius r(0)=rmin (r(1)=rmax) corresponds to the minimum (maximum) geometry.

We define a toleranced protein by a collection of toleranced balls, and an toleranced assembly by a collection of toleranced proteins. Since such an assembly has a continuum of possible geometries, we

*Extended abstract of the work to be presented at the `ISMB Satellite Meeting on Structural Bioinformatics and Computational Biophysics`, StoCkholm, 27-28 june 2009; see http://www.ebi.ac.uk/ rafi/3dsig09/Home.html

†INRIA Sophia-Antipolis-Méditerranée; Algorithms-Biology-Structure; Frederic.Cazals@sophia.inria.fr

‡INRIA Sophia-Antipolis-Méditerranée; Algorithms-Biology-Structure; Tom.Dreyfus@sopha.inria.fr

wish to select its stable features. To do so, imagine that all balls are grown simultaneously by letting the parameter l span the interval [0,1]. Typically, balls which are disjoint at l=0 merge at some point. Along the growth process, we track the evolution of the connected components (c.c.) of the union of growing balls. More precisely, a c.c. has a birth date lb (when two or more balls merge) and a death date ld (the c.c. merges with another one), so that a quantitative measure of the stability of the c.c. is given by ld-lb. Those which are long lived are the stable features of the assembly. An overall view of the growth process is encoded in a special graph called a Hasse diagram, which is a forest of trees, the length of a branch of a tree being precisely the quantity ld-lb of some c.c., that is, of a complex.

Technically speaking, the growth process just described is associated to a so-called multiplicatively-additively weighted Voronoi diagrams–CW-diagram for short. As opposed to the Voronoi diagrams used so far in molecular modeling, which are power i.e. affine diagrams, the bissectors of a CW-diagram are in general degree four algebraic surface patches. Since we are not aware of any algorithm able to construct such a diagram, we track the variation of the c.c. by probing this Voronoi diagram with a power diagram. The probing process consists of hunting the value of parameter l yielding a change in the number of c.c.

3 Application to uncertain assemblies reconstructed from tap data

The construction just sketched can be used to disambiguate TAP data. Assume that we are given a toleranced assembly, consisting say of n toleranced proteins (instances in the sequel) of p different types (types in the sequel). For example, for the NPC, one has n=456 and p=30. Consider now a set Q of q types (q_i=p) seen in a TAP experiment. By running the previous machinery on balls of the prescribed types in the toleranced assembly, we can state whether the set Q corresponds to an isolated complex, or to a mixture of (stable) complexes.

4 A case study: the nuclear pore complex

In [A07a,A07b], 1000 optimized structures of the NPC have been reconstructed from various experimental data including ultra-centrifugation data, cryo-EM, immuno-EM, overlay assays, and TAP data. These plausible structures have also been merged to define an average structure, which is of qualitative interest. Indeed, quoting the authors: "Our map is sufficient to determine the relative positions within NPC; we do not interpret features smaller than this precision" [A07a].

Starting from this mean structure, we show how to build a toleranced model of the NPC. Running our characterization of stable structures on this model provides a quantitative answer for deciding whether TAP data correspond to single complex or a mixture of complexes. This work has been carried out for the 83 pullouts of [A07a]. The example of composite #14, which contains Nup84 and Nup145C, is provided on Fig 1. The left panel shows a top view of a mixture of 16 complexes (8 on each side on the symmetry plane of the NPC), while the right panel provided the aforementioned Hasse diagram.

5 Conclusion

This work presents a method allowing one to make quantitative assessments for uncertain / toleranced models. In particular, it allows a structural interpretation of TAP data in conjunction with structural models of low to intermediate resolution. We believe this is an important step to build mechanistic models of large assemblies, so as to start investigating their dynamics.

REFERENCES

[G06] A-C. Gavin et al, Nature, 440, 2006.
[A07a] F. Alber et Al., Nature, 450, pp695-701, 2007.
[A07b] F. Alber et All, Nature, 450, pp683-694, 2007.
[A08] F. Alber et al., Ann. Rev. Biochem., 77, 2008.

Figure 1 Instantiation of composite #14, and stability assessment for its complexes. See text for details.