

Uso de Análise de Componentes Principais no estudo da heterogeneidade conformacional de complexos macromoleculares

Dr. Rodrigo Villares Portugal – Pesquisador Responsável

Laboratório Nacional de Nanotecnologia (LNNano)

1. Introdução

A visualização de complexos biológicos macromoleculares por técnicas de microscopia eletrônica é feita há várias décadas. No fim dos anos 50, a utilização de sais de metal pesado para o preparo das amostras mostrou-se uma técnica efetiva para melhoria do contraste apresentado por imagens de macromoléculas biológicas em solução. Embora tenha introduzido um significativo avanço, essa técnica ainda apresentava sérias limitações, entre elas a possibilidade de deformação estrutural da molécula e a exposição da amostra ao ambiente de vácuo do interior do microscópio. Com o desenvolvimento da técnica de preparação de amostras em gelo amorfo por Jacques Dubochet e colaboradores [Adrian 1984], o uso da microscopia como ferramenta de biologia molecular estrutural atingiu um novo patamar. O congelamento rápido da amostra evita a formação de cristais de gelo, preservando a integridade estrutural da molécula, mantendo-a da mesma maneira que se encontra em solução. O atual sucesso da microscopia eletrônica na biologia molecular estrutural está fortemente relacionado ao desenvolvimento dessa técnica. Com o surgimento dos métodos de análise de Partícula Única, e o recente aumento do poder computacional disponível, a Crio-Microscopia Eletrônica de Partícula Única se firmou como uma importante técnica de biologia molecular estrutural. Sua relevância vem sendo mostrada através do crescente número de estruturas depositadas no EM Data Bank - PDBe (www.ebi.ac.uk/pdbe) e da participação frequente em trabalhos de impacto em biologia molecular.

Amostras biológicas são bastante sensíveis ao feixe de elétrons do microscópio eletrônico, sendo necessário utilizar baixas doses de elétrons na obtenção das imagens. Estas condições degradam severamente a relação sinal-ruído das imagens adquiridas, tornando-as imagens tipicamente ruidosas, o que dificulta largamente seu processamento. A obtenção de dados é feita na ordem de dezenas ou centenas de milhares de imagens da macromolécula de interesse. Essas imagens são combinadas, em um processo iterativo, para obtenção de um modelo estrutural da molécula.

Até poucos anos atrás, a crio-ME de partícula única, apresentava algumas restrições, que aos poucos estão sendo superadas. O tamanho das moléculas passíveis de análise era tipicamente da ordem de MDa, não sendo possível resolver estruturas de tamanho inferior à várias centenas de kDa. As estruturas obtidas pela técnica de partícula única possuíam resolução de alguns nanômetros, tornando necessária a combinação

de técnicas de alta resolução para qualquer tipo de análise. Além disso, amostras heterogêneas inviabilizavam o tratamento de dados ou introduziam uma séria diminuição de resolução, em função da diversidade de estados presentes.

Esse cenário tem se alterado nos últimos anos. O tamanho dos complexos com estruturas resolvidas por crio microscopia eletrônica (crio-ME) de partícula única é hoje limitado a poucas centenas de kDa e a resolução obtida para complexos estáveis é tipicamente inferior a 1 nm. Recentemente, Zhou e colaboradores mostraram ser possível a obtenção do modelo atômico de uma proteína por crio-ME, com resolução de 3.3Å, sendo a primeira reconstrução *ab-initio* de estrutura atômica por crio-ME [Zhang 2010]. Mais importante, a possibilidade de se trabalhar com amostras heterogêneas é hoje uma realidade, sendo este um importante diferencial com relação a outras técnicas de biologia estrutural. A flexibilidade das macromoléculas, ou a existência de diferentes complexos, faz com que uma mesma amostra possa ter diferentes estruturas da macromolécula presentes. Sendo assim, a existência de diferentes estados conformacionais, ou estruturas, está ligada à dinâmica da proteína, podendo também estar associada à ligação de outras moléculas ou peptídeos. Várias metodologias vêm sendo propostas para se lidar com a necessidade de obter estruturas de dois ou mais complexos existentes na mesma amostra, sem a necessidade de separá-los para a aquisição de dados [Klaholz 2004; Penczek 2006; Leschziner 2007; van Heel 2012, Scheres 2012].

A proposta apresentada por Penczek e cols. (2006) relaciona a variância ao longo do modelo da macromolécula com a existência de uma ou mais estruturas combinadas no mesmo modelo. A figura 1 mostra um exemplo onde uma região de alta variância é identificada, indicando a existência de duas estruturas combinadas, uma com o EF-G e outra sem.

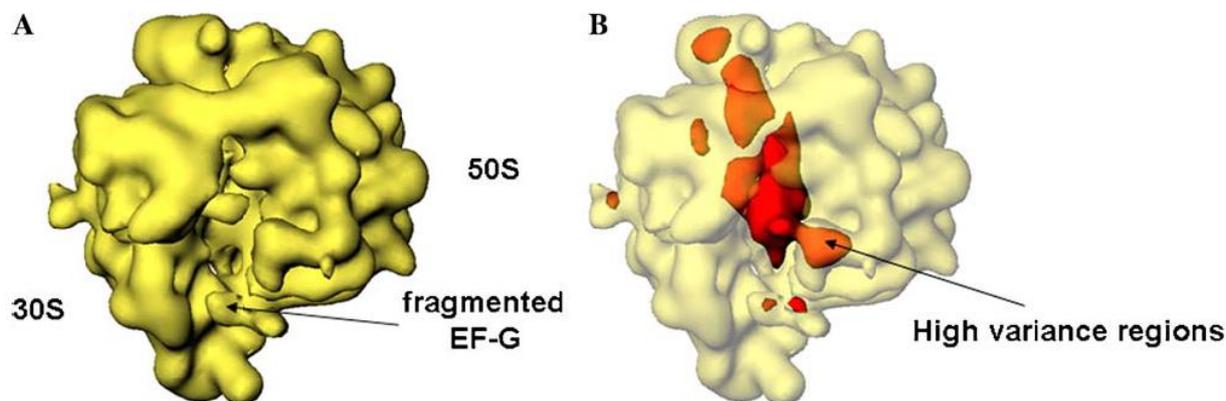


Figura 1. Análise da variância e resultados de classificação de uma reconstrução de ribossomo 70S de *E. coli*. (A) Reconstrução utilizando o conjunto completo de reprojeções. (B) Regiões de alta variância (vermelho) no complexo mostrado em (A). As regiões de alta variância estão associadas a diferenças na estrutura, no caso, a presença do EF-G (extraído de [Penczek 2006]).

Outras propostas de análise podem ser avaliadas no sentido de identificar a presença de uma ou mais estruturas nos modelos. Em particular, nos interessa avaliar o uso de Análise de Componentes Principais (Principal Component Analysis – PCA) de modelos 3D nesse contexto [van Heel 2009]. Em 2004, Klaholz e cols. prepuseram o uso de PCA para a separação de imagens oriundas de diferentes estruturas. Além do PCA de imagens, o PCA de estruturas 3D também vem sendo utilizado como ferramenta para análise de heterogeneidade conformacional.

2. Objetivo

Este projeto tem como objetivo o estudo do uso de Análise de Componentes Principais (PCA) de modelos (3D) para identificação de regiões de alta variância, e potencialmente, regiões flexíveis ou presença de outras moléculas.

3. Metodologia

Serão utilizadas diferentes estruturas de uma macromolécula conhecida (com PDB depositado), tal como o ribossomo. Estas estruturas serão utilizadas para gerar imagens de projeção da macromolécula, da mesma forma como são obtidas por microscopia eletrônica de transmissão. Tendo um conjunto de imagens, de duas ou mais estruturas, em diversas orientações, serão construídos uma quantidade suficiente de modelos para realizar a análise de componentes principais. Os resultados da análise serão avaliados buscando-se identificar relações entre a heterogeneidade conformacional, as regiões de maior variância e os resultados do PCA. Para as análises será utilizado o software IMAGIC [van Heel 2012].

Referências

- Adrian M, Dubochet J, Lepault J, McDowell AW, *Nature* **308** (1984) 32-36.
Klaholz BP, Myasnikov AG & van Heel M, *Nature* **427** (2004) 862-865 (especially: Supp. Info).
Leschziner AE & Nogales E, *Annu. Rev. Biophys. Biomol. Struct.* **36** (2007) 43-62.
Penczek PA, Frank J, Spahn CMT, *J. Struct. Biol.* **154** (2006) 184-194.
Scheres SHW. *J. Mol. Biol.* **415** (2012) 406-418.
Van Heel M, Portugal R, Schatz M, In: *Electron Microscopy in Life Science – An Electronic Textbook* (2009).
Van Heel M, Portugal R, Rohou A, Linnemayr, C, Bebeacua C, Schmidt R, Grant T, Schatz M, *International Tables for Crystallography, Vol. F.* (2012).
Zhang X, Jin L, Fang Q, Hui WH, Zhou ZH, *Cell* **141** (2010) 472-482.