Enhancing Evolutionary Couplings with Deep Convolutional Neural Networks

Graphical Abstract

Train on set of inferred couplings and known structures:

Co-evolution + PDB structure → Learn parameters

DeepContact convolutional neural network (CNN):

Co-evolution map from sequence → Run couplings through CNN → DeepContact contact map

Highlights

- Deep learning improves co-evolution-based protein residue-residue contact prediction
- Protein structure space constrains contact map space
- Recurrent co-evolutionary motifs appear across protein structures and families
- Models allow probabilistic interpretation of evolutionary couplings

Authors

Yang Liu, Perry Palmedo, Qing Ye, Bonnie Berger, Jian Peng

Correspondence

bab@mit.edu (B.B.), jianpeng@illinois.edu (J.P.)

In Brief

Many protein structures of interest remain out of reach for both computational prediction and experimental determination. DeepContact learns patterns of co-evolution across thousands of experimentally determined structures, identifying conserved local motifs and leveraging this information to improve protein residue-residue contact predictions. DeepContact extracts additional information from the evolutionary couplings using its knowledge of co-evolution and structural space, while also converting coupling scores into probabilities that are comparable across protein sequences and alignments.
Enhancing Evolutionary Couplings with Deep Convolutional Neural Networks

Yang Liu, Perry Palmedo, Qing Ye, Bonnie Berger, Jian Peng

INTRODUCTION

Protein structure and function are by nature intertwined, with structure, or structural properties, playing a large role in defining function. As such, ever since the X-ray structure of lysozyme led to the elucidation of its mechanism of catalytic action, determining protein structure has been one of the most important challenges in biology (Phillips, 1966; Phillips, 1967). In parallel, the many obstacles to experimental structure determination, computational prediction of protein structure remains one of the longest-standing challenges in computational biology (Moult et al., 2014, 2016). Existing approaches to protein structure prediction can be categorized into two types: template-based modeling and template-free modeling. With the requirement of a homologous structure, template-based methods are often not applicable to structure prediction tasks of interest, including orphan proteins, and thus for many novel proteins the field has turned to template-free, or de novo, folding approaches that predict 3D structures using the query sequence alone (Zhang, 2008; Xu and Zhang, 2012). While these approaches work reasonably well for smaller proteins, they have generally required further difficult-to-obtain experimental data for larger proteins (Bradley et al., 2005; Moult et al., 2014, 2016).

Recent computational advances, together with the availability of large protein sequence databases, have enabled us to exploit rich evolutionary information encoded within multiple sequence alignments (MSAs) to assist traditional protein structure prediction approaches. Notably, evolutionary coupling analysis methods, such as direct-coupling analysis, GREMLIN, (meta-) PSICOV, and EVFold, take an MSA as input and predict residue-residue contacts by learning an inherent graphical model structure that incorporates pairwise terms to capture evolutionary constraints among residues (Ekeberg et al., 2013; Morcos et al., 2011; Kamisetty et al., 2013; Jones et al., 2012, 2015; Marks et al., 2011; Kaján et al., 2014). Several tools (including Rosetta) have successfully incorporated evolutionary couplings into their pipelines as distance restraints to significantly improve predictions, particularly for proteins that have proven challenging using traditional approaches (Ovchinnikov et al., 2014, 2015; Weinreb et al., 2016). In addition, evolutionary coupling-based methods have been successfully applied to protein complex assembly and interactions, disordered region structure prediction, RNA structure prediction, and mutagenesis analysis (Ovchinnikov et al., 2014; Uguzzoni et al., 2017; Hopf et al., 2014, 2017; Toth-Petroczy et al., 2016; De Leonardis et al., 2015; Weinreb et al., 2016).

Despite these advances, state-of-the-art evolutionary coupling approaches still have several major failings that limit their applicability. First, they require large, high-quality MSAs and often generate sparse or poor contact predictions for proteins with less robust MSAs (Moult et al., 2016). Second, the units of evolutionary couplings are arbitrary; while there have been recent attempts to define significant couplings, for the most part users are left to decide how many couplings to take
دریافت فوری متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات