Philipp Heuser, Victor Lamzin
European **M**olecular **B**iology **L**aboratory, c/o DESY, Notkestrasse 85, 22607 Hamburg, Germany
[philipp.heuser | victor]@embl-hamburg.de

EMBL

Pattern recognition with moment invariants for interpretation of very low resolution macromolecular density maps

Due to the growing amount of low-resolution structural data for macromolecular protein complexes from Electron Microscopy, X-ray Crystallography, XFEL and other methods, novel approaches for the interpretation of such data are urgently required. Here we present a method based on pattern recognition using third-order moment invariants, to automatically place known domain structures in a low-resolution density map without user intervention. The approach does not require prior segmentation of the map or the presence of internal symmetry.

INTRODUCTION

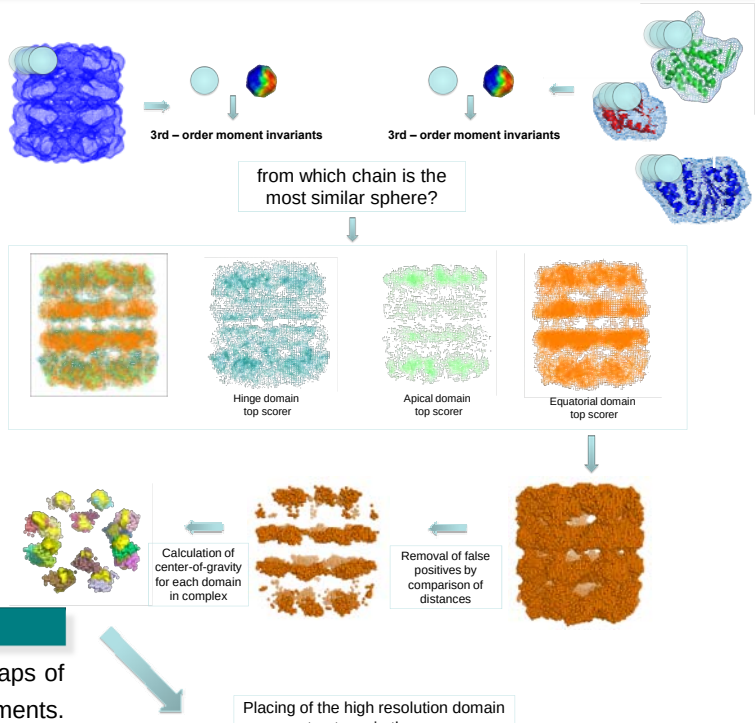
Current development of methods for **solving macromolecular structures** is largely focused on high-resolution data. However, many of the most challenging questions, which have to be addressed by modern structural biology, require an analysis of large macromolecular **complexes**, for which only in rare cases high-resolution data can be obtained.

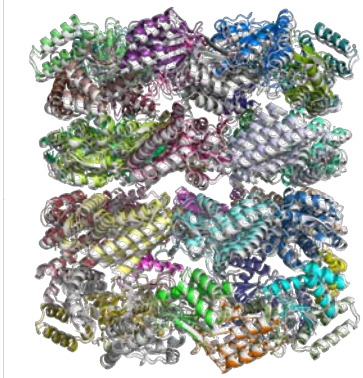
The interpretation of **very low-resolution data** (>10Å) usually starts with the segmentation of the map (e.g., Watershed algorithm), which does not always give satisfactory results. Overall, **fitting of known structures** currently requires a lot of human expert knowledge and interaction. Instead, an **automated objective procedure** is highly desirable.

METHOD

We use **3rd-order moment invariants** to identify regions in density maps of macromolecular complexes that match known structures or their fragments. False positives are eliminated using **difference distance matrices**. As a result, the structures of the fragments are **placed into the map**.

Third-order moment invariants give a concise but comprehensive description of 3D objects, providing convenient means for **fast searches through a large amount of 3D data**.





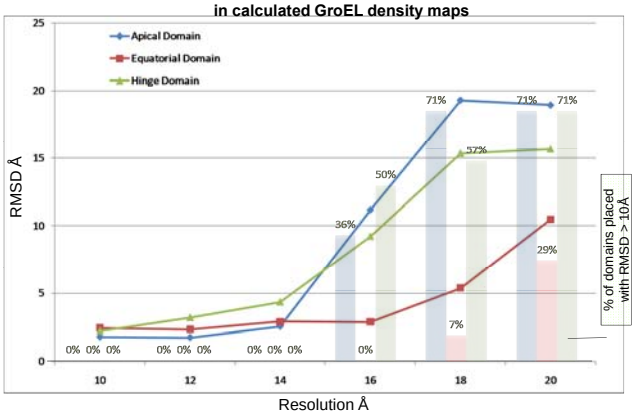
RMSD:
Average: 2.16Å
min: 0.65Å max: 3.88Å

Apical domain: 1.79Å
Equatorial domain: 2.44Å
Hinge domain: 2.24Å

RESULTS

The method has been tested on calculated density maps for large macromolecular complex of GroEL with high-resolution limit between 10 and 20Å. The individual domains were placed in the low-resolution density maps with an average **RMSD on Cα atoms of 2.2Å at 10Å resolution**.

Average RMSD of placed domains in calculated GroEL density maps



CONCLUSION

For the last decade there have been many attempts to develop reliable 3D map segmentation algorithms with varying success, in order to reduce the complexity of the challenging task of low-resolution density map interpretation. The method presented here does **not require a map segmentation step** and provides accurate results **without human interaction** in reasonable time, due to the use of sophisticated pattern recognition algorithms. Implementation of real-space refinement procedures is expected to improve the results even further.