Automated shape-based clustering of 3D immunoglobulin protein structures in chronic lymphocytic leukemia

**Background:** Although the etiology of chronic lymphocytic leukemia (CLL), the most common type of adult leukemia, is still unclear, strong evidence implicates antigen involvement in disease ontogeny and evolution. Primary and 3D structure analysis has been utilised in order to discover indications of antigenic pressure. The latter has been mostly based on the 3D models of the clonotypic B cell receptor immunoglobulin (BcR IG) amino acid sequences. Therefore, their accuracy is directly dependent on the quality of the model construction algorithms and the specific methods used to compare the ensuing models. Thus far, reliable and robust methods that can group the IG 3D models based on their structural characteristics are missing.

**Results:** Here we propose a novel method for clustering a set of proteins based on their 3D structure focusing on 3D structures of BcR IG from a large series of patients with CLL. The method combines techniques from the areas of bioinformatics, 3D object recognition and machine learning. The clustering procedure is based on the extraction of 3D descriptors, encoding various properties of the local and global geometrical structure of the proteins. The descriptors are extracted from aligned pairs of proteins. A combination of individual 3D descriptors is also used as an additional method. The comparison of the automatically generated clusters to manual annotation by experts shows an increased accuracy when using the 3D descriptors compared to plain bioinformatics-based comparison. The accuracy is increased even more when using the combination of 3D descriptors.

**Conclusions:** The experimental results verify that the use of 3D descriptors commonly used for 3D object recognition can be effectively applied to distinguishing structural differences of proteins. The proposed approach can be applied to provide hints for the existence of structural groups in a large set of unannotated BcR IG protein files in both CLL and, by logical extension, other contexts where it is relevant to characterize BcR IG structural similarity. The method does not present any limitations in application and can be extended to other types of proteins.

**General information**
- **Publication status:** Published
- **Organisations:** Department of Applied Mathematics and Computer Science , Immunoinformatics and Machine Learning, Department of Bio and Health Informatics, Center For Research And Technology - Hellas, North Shore-Long Island Jewish Health System, Masaryk University, Vita-Salute San Raffaele University, Carlsberg Research Center, Ionian University, Nikea General Hospital, Papanikolaou Hospital, Technical University of Denmark
- **Corresponding author:** Polychronidou, E.
- **Number of pages:** 15
- **Publication date:** 2018
- **Peer-reviewed:** Yes

**Publication information**
- **Journal:** BMC Bioinformatics
- **Volume:** 19
- **Issue number:** Suppl. 14
- **Article number:** 414
- **ISSN (Print):** 1471-2105
- **Ratings:**
  - BFI (2018): BFI-level 1
  - Web of Science (2018): Indexed yes
- **Original language:** English
- **Keywords:** 3D protein descriptors, CLL protein clustering, Descriptor fusion
- **Electronic versions:**
  - 10.1186_2Fs12859_018_2381_1.pdf
  - DOIs:
    - 10.1186/s12859-018-2381-1
- **Source:** Scopus
- **Source-ID:** 85056744630
- **Research output:** Contribution to journal → Conference article – Annual report year: 2018 → Research → peer-review