Databases and ontologies

PHI-DAC: protein homology database through dihedral angle conservation

Noa Maatuk, Yitav Glantz-Gashai, Maya Rotman, Meirav Baydany, Gennadiy Fonar, Amir Shechvitz, Karin Shemer, Aviva Peleg, Eli Reuveni and Abraham O. Samson*

Faculty of Medicine in the Galilee, Bar Ilan University, 13100 Safed, Israel

ABSTRACT
Finding related conformations in the Protein Data Bank is essential in many areas of bioscience. To assist this task, we designed a dihedral angle database for searching protein segment homologs. The search engine relies on encoding of the protein coordinates into text characters representing amino acid sequence, \( \varphi \) and \( \psi \) dihedral angles. The search engine is advantageous owing to its high speed and interactive nature and is expected to assist scientists in discovering conformation homologs and evolutionary kinship. The search engine is fast, with query times lasting a few seconds, and freely available at http://tarshish.md.biu.ac.il/~samsona

Here we describe an online search engine that rapidly finds protein segments based on amino acid sequence and \( \varphi \) and \( \psi \) dihedral angles. The search engine, named Protein Homology through Dihedral Angle Constraints (PHI-DAC), is advantageous owing to its speed, generality and simplicity. It is expected to be helpful to the scientific community by easing the identification of conformation homologs and distilling useful information from the PDB.

2 METHODS
Protein structure representation. To describe the 3D conformation of a protein we use \( \varphi \) and \( \psi \) dihedral angles. These angles have the advantage of being invariant to translation and rotation of the protein structure in a coordinate system. The angles are calculated using DSSP (Kabsch and Sander, 1983) and encoded into a structural alphabet (Levitt, 1992) represented by text characters from 0 to 359°. This transforms the information of the 3D backbone conformation from all protein structures contained in the PDB into sequences of text that are easily searchable.

Construction of the PHI-DAC database. All PDB files containing protein structure are handled as separate entries and included in PHI-DAC. As of January 2014, a total of \(~96,000\) files describe protein structures. Computation of the PHI-DAC database, given the \( \varphi \) and \( \psi \) dihedral angles and amino acid sequences, takes \( \sim\)1 min on a standard PC (2.7 GHz). The size of the database is 174 MB, and it can be held in working memory, making all calculations extremely fast. The search engine is hosted on a 2.7 GHz PC with 2 GB random access memory operating under Linux. Update of the PHI-DAC database is expected monthly.

3 RESULTS
PHI-DAC database. In Table 1, an example entry of the PHI-DAC database is shown. The entry is composed of four pipe delimited fields. The first and second fields contain the PDB identification (ID) and the amino acid sequence, respectively. Chain IDs are referenced to the DSSP files, and cysteines are in lowercase letters. Duplicate lowercase letters form disulfide bridges, ‘a’ with ‘a’, ‘b’ with ‘b’, etc. Exclamation points denote a chain break or a new polypeptide chain. The third and fourth fields contain the \( \varphi \) and \( \psi \) dihedral angles encoded in double characters. The first character represents units of 10 \( (A = 0, B = 10 \ldots S = 180, \text{and } a = -0, b = -10 \ldots) \) and the second character represents units of 1. Thus, ‘A0’ is \( 0° \), ‘A1’ is \( +1° \), ‘A2’ is \( +2° \), ‘S0’ is \( +180° \) (or \( -180° \)), ‘a1’ is \( -1° \) (or \( 359° \)).

*To whom correspondence should be addressed.
and so on. This protein backbone representation is concise and accurate. The database is available for download.

**Querying PHI-DAC.** To perform a search for similar protein backbone conformations in the PDB, the web interface offers the following search parameters: PDB ID, chain, and amino acid sequence. The PDB IDs are hyperlinked to the respective PDB file. The results are classified according to the dihedral angle root-mean-square deviation (RMSD) between the query segment and the result segment (low to high).

**Example queries.** Two example queries are readily accessible on the Web site. In the first example query the coordinates of a small bent helix in PDB ID 9XIM (residues 121–127) are used. Searching with the ϕ and ψ dihedral angles by using a tolerance of ±3° and without conserving the amino acid sequence leads to 14 matching segments belonging to 14 different PDB entries. In the second example query, the coordinates of the β-hairpin segment 7–10 of chain A of PDB ID 1L4W is used. Querying with the ϕ and ψ dihedral angles by using a tolerance of ±8° and without conserving the amino acid sequence leads to 13 matching segments belonging to 13 different PDB entries.

**Local versus global homology.** PHI-DAC is based on local dihedral angle homology and is not designed to detect global spatial similarities. As such, PHI-DAC does not detect similarity between large SCOP or CATH folds with perfect superimposition in the core and major differences in the loops. For spatial superimposition, SCOP, CATH and Dali should be used.

**4 CONCLUSION**

Our method of encoding the backbone ϕ and ψ dihedral angles into text characters has proven to be a fast solution for answering queries about local structural similarities in the PDB. Compared with SPASM, PHI-DAC shows similar results, while being much faster and addressing a different question compared with Dali. Therefore, compared with Dali, we consider PHI-DAC to be a valuable tool for the fast detection of protein segments based on dihedral angle conservation.

**ACKNOWLEDGEMENT**

We thank Prof. Haim Breitbart for helpful comments.

**Funding:** This research was supported by the Katz Foundation and a Marie-Curie CIG grant 322113 (to A.O.S.).

**Conflict of interest:** none declared.

**REFERENCES**


