

Poster presentation

## **PAUL: protein structural alignment using integer linear programming and Lagrangian relaxation**

Inken Wohlers\*<sup>1</sup>, Lars Petzold<sup>2</sup>, Francisco S Domingues<sup>3</sup>  
and Gunnar W Klau<sup>1</sup>

Address: <sup>1</sup>Life Sciences Group, Centrum Wiskunde & Informatica, Science Park 123, 1098 XG Amsterdam, the Netherlands, <sup>2</sup>Mathematics in Life Sciences Group, Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany and <sup>3</sup>Computational Biology and Applied Algorithmics Group, Max-Planck-Institut für Informatik, 66123 Saarbrücken, Germany

E-mail: Inken Wohlers\* - I.Wohlers@cw.i.nl

\*Corresponding author

from Fifth International Society for Computational Biology (ISCB) Student Council Symposium  
Stockholm, Sweden 27 June 2009

Published: 19 October 2009

*BMC Bioinformatics* 2009, **10**(Suppl 13):P2 doi: 10.1186/1471-2105-10-S13-P2

This article is available from: <http://www.biomedcentral.com/1471-2105/10/S13/P2>

© 2009 Wohlers et al; licensee BioMed Central Ltd.

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### **Background**

Protein structural alignment determines the three-dimensional superposition of protein structures by means of aligning the protein's residues. It is a basic method for identifying proteins of related structure or common evolutionary origin and for measuring three-dimensional similarity. Applications are for instance the search for proteins with similar biological function or the classification of proteins based on their structural features.

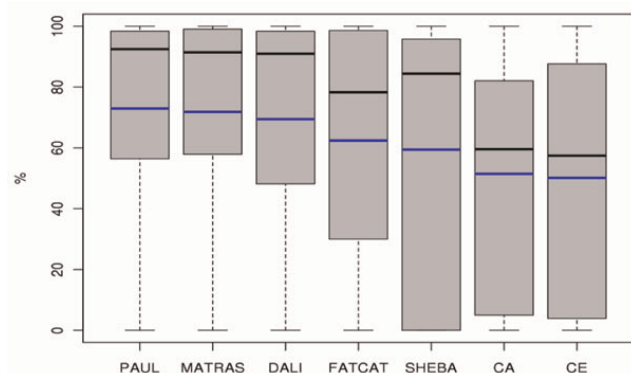
### **Methods**

We present a structural alignment approach that computes an alignment based on the protein's inter-residue distances. Building upon work for the alignment of protein contact maps by Caprara et al. [1], we use these distances to formulate the problem as an integer linear program which is subsequently solved using Lagrangian relaxation. One advantage of the integer linear programming formulation over heuristic methods is that we compute in many cases demonstrably optimal alignments. The bottleneck of the integer linear programming approach is its computational complexity which does not allow to incorporate all inter-

residue distances in the problem description. On that account we select and score inter-residue distances efficiently. We develop and optimize a scoring function inspired by Holm and Sander. [2] using a set of 200 pairwise HOMSTRAD [3] alignments with a sequence identity of less than 35%. Subsequently, we use this scoring function to assess the performance of PAUL on the more challenging SISY data set of 130 alignments [4,5] – on this data set we compare PAUL alignments to alignments computed by MATRAS [6], DALI [2], FATCAT [7], SHEBA [8], CA [9] and CE [10].

### **Results and conclusion**

Our novel, non-heuristic structural alignment algorithm is flexible and mathematically sound. On the SISY data set PAUL alignments show higher mean and median alignment accuracies than all other methods (see Figure 1). In more than 30% of the cases, PAUL is the most accurate method. PAUL is thus competitive to other state-of-the-art algorithms and a beneficial tool for high-quality pairwise structural alignment.



**Figure 1**  
**Box-and-whisker plots of the distributions of the percentages of alignment accuracies for the SISY set for PAUL, MATRAS, DALI, FATCAT, SHEBA, CA and CE. Additionally, the average alignment accuracies are denoted in blue.**

## References

1. Caprara A, Carr R, Istrail S, Lancia G and Walenz B: **1001 optimal PDB structure alignments: integer programming methods for finding the maximum contact map overlap.** *J Comput Biol* 2004, **11(1)**:27–52.
2. Holm L and Sander C: **Protein structure comparison by alignment of distance matrices.** *J Mol Biol* 1993, **233(1)**: 123–138.
3. Mizuguchi K, Deane CM, Blundell TL and Overington JP: **Homstrad: a database of protein structure alignments for homologous families.** *Protein Sci* 1998, **7(11)**:2469–2471.
4. Andreeva A, Prlic A, Hubbard TJ and Murzin AG: **Sisyphus-structural alignments for proteins with non-trivial relationships.** *Nucleic Acids Res* 2007, **35 Database**: 253–259.
5. Mayr G, Domingues FS and Lackner P: **Comparative analysis of protein structure alignments.** *BMC Struct Biol* 2007, **7**:50–50.
6. Kawabata T: **Matras: A program for protein 3d structure comparison.** *Nucleic Acids Res* 2003, **31(13)**:3367–3369.
7. Ye Y and Godzik A: **Flexible structure alignment by chaining aligned fragment pairs allowing twists.** *Bioinformatics* 2003, **19(Suppl 2)**:ii246–ii255.
8. Jung J and Lee B: **Protein structure alignment using environmental profiles.** *Protein Eng* 2000, **13(8)**:535–543.
9. Bachar O, Fischer D, Nussinov R and Wolfson H: **A computer vision based technique for 3-D sequence-independent structural comparison of proteins.** *Protein Eng* 1993, **6(3)**: 279–288.
10. Shindyalov IN and Bourne PE: **Protein structure alignment by incremental combinatorial extension (CE) of the optimal path.** *Protein Eng* 1998, **11(9)**:739–747.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

