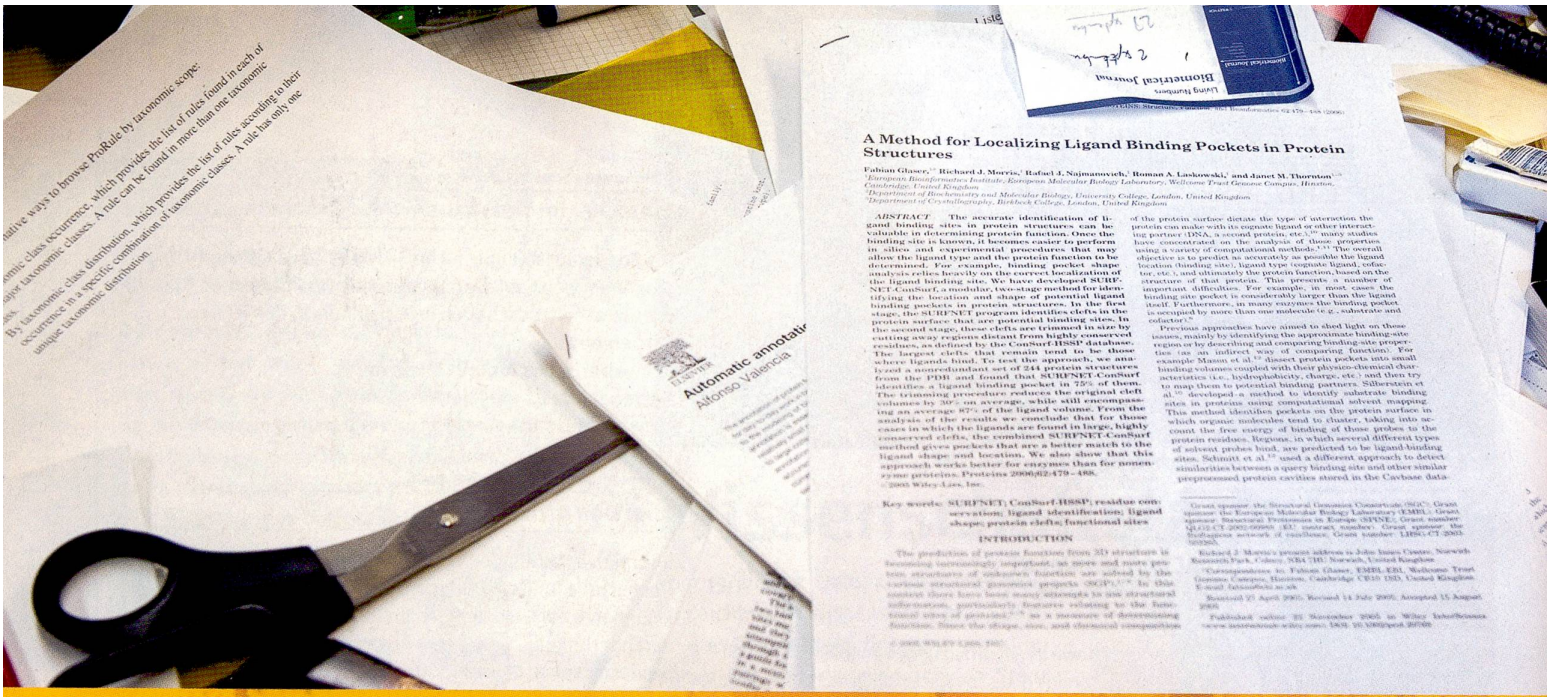

A Method for Localizing Ligand Binding Pockets in Protein Structures

Fabian Glaser,[1]* Richard J. Morris,[2] Rafael J. Najmanovich,[1] Roman A. Laskowski,[1] and Janet M. Thornton[1,3]
[1]European Bioinformatics Institute, European Molecular Biology Laboratory, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom
[2]Department of Biochemistry and Molecular Biology, University College, London, United Kingdom
[3]Department of Crystallography, Birkbeck College, London, United Kingdom

ABSTRACT **The accurate identification of ligand binding sites in protein structures can be valuable in determining protein function. Once the binding site is known, it becomes easier to perform in silico and experimental procedures that may allow the ligand type and the protein function to be determined. For example, binding pocket shape analysis relies heavily on the correct localization of the ligand binding site. We have developed SURFNET-ConSurf, a modular, two-stage method for identifying the location and shape of potential ligand binding pockets in protein structures. In the first stage, the SURFNET program identifies clefts in the protein surface that are potential binding sites. In the second stage, these clefts are trimmed in size by cutting away regions distant from highly conserved residues, as defined by the ConSurf-HSSP database. The largest clefts that remain tend to be those where ligands bind. To test the approach, we analyzed a nonredundant set of 244 protein structures from the PDB and found that SURFNET-ConSurf identifies a ligand binding pocket in 75% of them. The trimming procedure reduces the original cleft volumes by 30% on average, while still encompassing an average 87% of the ligand volume. From the analysis of the results we conclude that for those cases in which the ligands are found in large, highly conserved clefts, the combined SURFNET-ConSurf method gives pockets that are a better match to the ligand shape and location. We also show that this approach works better for enzymes than for nonenzyme proteins. Proteins 2006;62:479–488.**
© 2005 Wiley-Liss, Inc.

Key words: SURFNET; ConSurf-HSSP; residue conservation; ligand identification; ligand shape; protein clefts; functional sites

INTRODUCTION

The prediction of protein function from 3D structure is becoming increasingly important, as more and more protein structures of unknown function are solved by the various structural genomics projects (SGPs).[1–3] In this context there have been many attempts to use structural information, particularly features relating to the functional sites of proteins,[4–6] as a measure of determining function. Since the shape, size, and chemical composition

Grant sponsor: the Structural Genomics Consortium (SGC); Grant sponsor: the European Molecular Biology Laboratory (EMBL); Grant sponsor: Structural Proteomics in Europe (SPINE); Grant number: QLG2-CT-2002-00993 (EU contract number); Grant sponsor: the Intelligence network of excellence; Grant number: LHSG-CT-2003-503265.

Richard J. Morris's present address is John Innes Centre, Norwich Research Park, Colney, NR4 7UH, Norwich, United Kingdom.

*Correspondence to: Fabian Glaser, EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom. E-mail: fabian@ebi.ac.uk

Received 27 April 2005; Revised 14 July 2005; Accepted 15 August 2005

Published online 25 November 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20769