ABSTRACT
Summary: COPS predicts for all 20 naturally occurring amino acids whether the peptide bond in a protein is in cis or trans conformation. The algorithm is based only on secondary structure information of amino acid triplets without considering the amino acid sequence information. Conformation parameters are derived from solved 3D structures deposited in the PDB and led to propensities based on modified Chou–Fasman parameters. COPS analyses amino acid triplets taking only their respective secondary structure into consideration and upon application of a set of rules utilizing the conformation parameters, the N-terminal peptide bond conformation of the middle residue is predicted. COPS was tested on a random selection of protein datasets.
Availability: The COPS program and further information are freely available from the FMP website at http://www.fmp-berlin.de/nmr/cops
Contact: labudee@fmp-berlin.de

The majority of the peptide bond conformations in protein structures are found to be in trans (Ramachandran and Sasisekharan, 1968). For proline, the situation is different: here, the sterical differences between cis and trans are minimal. A survey by Stewart et al. (1990) found only 0.05% of all Xaa-nonPro, but 6.5% of all Xaa-Pro peptide bonds to occur in the cis conformation. MacArthur and Thornton (1991) found 5.5% cis for Xaa-Pro and Weiss et al. (1998) found in a much larger non-redundant set of 571 proteins 5.2% cis for Xaa-Pro and 0.03% cis for Xaa-nonPro.

The possible biological role of the cis/trans isomerization, especially for prolines, in protein folding, splicing, active transport through membranes, and energy reservoir is still a matter of some debate (Andreotti, 2003; Fischer and Aumuller, 2003).

The aim of this work was to develop an algorithm for the prediction of the cis and trans conformation of all naturally occurring amino acids in proteins. The new algorithm is based on an extension of the Chou–Fasman parameters (Chou and Fasman, 1974). This extension contains the separation between cis and trans conformation for the considered amino acid. These new parameters are called conformation parameters.

We calculate conformation parameters for each amino acid employing the correlation between the secondary structure information and the cis/trans conformation of known structures deposited in the PDB (http://www.rcsb.org/pdb). We considered only those proteins which contained at least one cis residue. The resulting set of 8584 proteins yielded 25 663 amino acids in cis and about 11 million amino acids in trans conformation.

Three sets of conformation parameters were calculated. The first set is described by the conformation parameters for the cis/trans propensity for an individual residue based solely on the secondary structure at this residue. The second and third sets use in addition the secondary structure at positions \(i - 1\) and \(i + 1\) as conditional propensities. These conditional propensities result from occurrences of a secondary structure element at the positions \(i - 1\) and \(i + 1\) respectively depending on the secondary structure at the position \(i\).

Based on our three sets of conformation parameters, we derive rules for the peptide bond conformation prediction between residues \(i - 1\) and \(i\) considering the secondary structure of the neighboring residues. Hence, for an investigated protein COPS analyses the secondary structure of amino acid triplets. The prediction if a residue of the investigated protein sequence is in the cis conformation is based on four rules, all of which have to be fulfilled otherwise trans is predicted. The first criteria is that the sum of the cis conformation parameters has to be greater than the sum of the trans conformation parameters. Secondly, the cis conformation parameter at position \(i\) has to be greater than the trans conformation parameter. Thirdly, the cis conformation parameter of both neighboring residues must be greater then zero. Fourthly, one value of the bordering residues must be greater than 1.

We tested our algorithm against randomly selected protein sequences of our database. The error rate varies depending on the amino acid which was predicted (details are presented on the webpage). We evaluated the performance of COPS by calculating the statistical significance, the statistical sensitivity, the positive and negative predictive values of the COPS prediction for all amino acids. Equations, description of the datasets and results are reported on our webpage. The new algorithm could be helpful in the structure determination process to check for possible cis conformations or in conjunction with secondary structure prediction tools to detect possible cis bonds in structurally unknown proteins.

In contrast to most other peptide bond conformation prediction methods which consider only prolines (Schubert et al., 2002; Frommel and Preissner, 1990) COPS can predict the cis/trans conformation for all amino acids. Weiss and Hilgenfeld (1999) proposed...
a method for all amino acids which relies on a completely solved structure.

COPS has been implemented in Tcl/Tk (8.3) to allow for full cross-
platform compatibility and has been successfully tested on SGI Irix,
Linux and Microsoft Windows. The program COPS and supporting
information are available at http://www.fmp-berlin.de/nmr/cops.

ACKNOWLEDGEMENTS
Support from the Forschungsinstitut für Molekulare Pharmakologie
is gratefully acknowledged. The work was supported by a grant of
the BMBF (01 GG 9812, Leitprojekt ‘Proteinsstrukturfabrik’). We are
grateful to K. Heuer for carefully reading the manuscript.

REFERENCES
Biochemistry, 42, 9515–9524.

Chou,P. and Fasman,G. (1974) Conformational parameters for amino acids in helical,
beta-sheet, and random coil regions calculated from proteins. Biochemistry, 13,
211–222.


protein structures on the basis of the amino acid sequence. FEBS Lett., 277,
159–163.


Ramachandran,G. and Sasisekharan,V. (1968) Conformation of polypeptides and

the prediction of Xaa-Pro peptide bond conformations in proteins based on 13C chemical


5, 676.