

PublisherInfo		
PublisherName	:	BioMed Central
PublisherLocation	:	London
PublisherImprintName	:	BioMed Central

SIFTing the effects of SNPs

ArticleInfo		
ArticleID	:	3925
ArticleDOI	:	10.1186/gb-2001-2-7-reports0019
ArticleCitationID	:	reports0019
ArticleSequenceNumber	:	19
ArticleCategory	:	Paper report
ArticleFirstPage	:	1
ArticleLastPage	:	4
ArticleHistory	:	RegistrationDate : 2001-6-1 Received : 2001-6-1 OnlineDate : 2001-6-27
ArticleCopyright	:	BioMed Central Ltd2001
ArticleGrants	:	

Reiner Veitia

Abstract

An *in silico* approach has been used to sort the effects on protein structure or function of amino-acid substitutions at conserved compared to non-conserved sites.

Significance and context

It is clear that not all sites in homologous proteins are conserved to the same extent. Those that are essential will be highly conserved (intolerant of change), whereas others that are less important for structure and function will be under less evolutionary constraint (tolerant of change). Here, Ng and Henikoff describe an algorithm, SIFT, a sequence homology-based method that sorts intolerant from tolerant amino-acid substitutions. By aligning multiple similar sequences and assessing the probability of substitution at any give position in the sequence, SIFT helps to assess the impact of an amino-acid replacement on the structure or function of a protein. This method might be useful in the following circumstances: during mutation screening when the status of a mutation suspected to be pathogenic cannot be formally shown (for example, in the absence of parental DNA); to assess the impact of amino-acid substitutions on fitness at a genomic scale; and in population genetics, to avoid using markers that may be undergoing selective pressure.

Key results

SIFT takes a query sequence and searches for similar sequences using well known tools (PSI-BLAST and MOTIF). Then, a multiple sequence alignment is obtained and the normalized probabilities for all possible substitutions at each position of the alignment are calculated (providing position-specific information). If the probability of the substitution is lower than a specified cutoff, the change is considered to be deleterious. The performance of SIFT was tested using three mutation data sets: the repressor of the lactose operon, LacI; the HIV-1 protease; and the bacteriophage T4 lysozyme. The prediction accuracy of SIFT is in the range of 60-80%, depending on the data set. In all cases, the performance of SIFT has been compared with the conclusions drawn from the look-up scoring matrix BLOSUM62 (Block substitution matrix), which is used, as are many others, to assess the significance of a protein sequence alignment (as in BLAST). BLOSUM62 helps to distinguish between a 'real' biological result and a sequence alignment obtained by chance. In BLOSUM, each possible amino-acid change is assigned a score, where positive scores will be associated with conservative changes and

negative scores with less conservative changes. Position-specific information is lost in the BLOSUM matrix, but is retained by SIFT, so SIFT outperforms BLOSUM62-derived conclusions.

Links

The algorithm can be downloaded from [SIFT](#) maintained by Pauline Ng. A report of another approach to assessing the effects of mutation on protein structure and function can be found on the [Genome Biology 2\(7\):reports](#)

Reporter's comments

SIFT relies solely on sequence homology and is suitable for automation. For the moment, the limiting step is the collection of 'homologous' sequences, but this problem will vanish as more genomic and cDNA sequences become available from the ongoing and new genome projects. The method is expected to perform best in the analysis of homologous proteins with conserved functions.

Table of links

[Genome Research](#)

[SIFT](#)

[Genome Biology 2\(7\):reports](#)

References

1. Ng PC, Henikoff S: Predicting deleterious amino acids substitutions. *Genome Res.* 2001, 11: 863-874. 1088-9051

