

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

## Finding motifs in protein sequences

ArticleInfo		
ArticleID	:	3655
ArticleDOI	:	10.1186/gb-2000-1-3-reports2052
ArticleCitationID	:	reports2052
ArticleSequenceNumber	:	24
ArticleCategory	:	Web report
ArticleFirstPage	:	1
ArticleLastPage	:	4
ArticleHistory	:	RegistrationDate : 2000-8-2 Received : 2000-8-2 OnlineDate : 2000-9-18
ArticleCopyright	:	BioMed Central Ltd2000
ArticleGrants	:	

ArticleContext	:	130591133
----------------	---	-----------

Todd Richmond

---

## Abstract

EMOTIF search is one of a set of four integrated bioinformatics resources at Stanford University devoted to constructing and searching for motifs in protein sequences.

## Content

EMOTIF search is one of a set of four integrated bioinformatics resources at Stanford University devoted to constructing and searching for motifs in protein sequences. EMOTIF search, the subject of this report, finds motifs in user-specified proteins; [EMOTIF maker](#) will construct motifs from multiple sequence alignments of protein; [EMOTIF scan](#) allows you to search for proteins that contain a motif that you specify; and [3MOTIF](#) allows you to find three-dimensional motifs using protein structures.

## Navigation

There is no difficulty in navigating the EMOTIF search site. The page has a box for pasting the protein sequence and two buttons: one to initiate the search, and another to clear the form. There are links in the upper corner of the page to all four of the related resources mentioned above. After submitting the search, the user is presented with a page of possible motif matches, in decreasing order of stringency. Each of the matches has a short description and a link to the complete motif description in the [BLOCKS](#) database. The location of the motif in the protein is specified and some of the sequence around it is displayed, and, when available, there is a link to the three-dimensional representation of the motif in 3MOTIF. You can also choose a link to find other proteins with the same motif, using EMOTIF scan.

## Reporter's comments

## Timeliness

There is no indication of when the site was last updated, or what version of each of the sequence databases is being searched.

## Best feature

The site is very simple to use, and the integration of the various resources is very useful. One can make a motif, search for proteins with the motif, and then determine if they, in turn, share any other motifs.

## Worst feature

Unfortunately, the results are of dubious use. Using one of my favorite proteins - a putative glycosyltransferase from *Arabidopsis* - one of the true conserved motifs was buried in a mess of false positives (though the page claims that no false positives are expected at that stringency). Worse, when I went to check on the description of the 'true hit' in the BLOCKS database using the supplied link, I received an error saying that no such BLOCK exists. When I used the link to initiate an EMOTIF scan, I was presented with a substantial list of matching proteins, from both SwissPROT and GenBank. But closer inspection revealed that a number of proteins that should have matched the same motif were not present. In fact, of the 22 known *Arabidopsis* proteins with this particular glycosyltransferase motif, not a single one was in the list - a very glaring omission. In the interests of fairness, I decided to test another protein: a multifunctional protein involved in beta-oxidation of fatty acids. There are several very clear domains in this protein, which match the PROSITE consensus sequences for these motifs. One domain was identified (in fact, 18 times), but the other domains were not. An EMOTIF scan with several of the motif matches again revealed an absence of any of the *Arabidopsis* sequences that contain these motifs. Although it is not stated anywhere on the site, it seems clear that only a subset of the protein database (or a very old version) is being searched.

When I tried to allow a single mismatch in the EMOTIF scan, thinking that perhaps a single amino-acid mismatch might cause some proteins to be omitted, I discovered that this feature is obviously broken. Instead of a short list of matching proteins with the protein motif highlighted, the search instead started spewing an incredible number of full-length protein sequences, without any highlighting or notation.

It should be noted that the EMOTIF site has undergone some revisions in the month since this report was written. The navigation has not changed and there still appear to be problems with the results - now it is more likely that no results will be returned than the user will be given spurious ones.

## Wish list

The site needs better documentation to let people know how the programs work and to state clearly the limitations of the tools. I searched through most of the site and the only help pages I could find were for the construction of EMOTIFs from multiple sequence alignments.

## Related websites

There is no indication of when the site was last updated, or what version of each of the sequence databases is being searched.

Two better sites for motif searches are the [BLOCKS](#) servers and the [PROSITE database of protein families and domains](#).

## Table of links

[EMOTIF search](#)

[EMOTIF maker](#)

[EMOTIF scan](#)

[3MOTIF](#)

[BLOCKS](#)

[PROSITE database of protein families and domains](#)

## References

1. [EMOTIF search](#).