BIOINFORMATICS APPLICATIONS NOTE Vol. 24 no. 18 2008, pages 2101–2102
doi:10.1093/bioinformatics/btn392

*Structural bioinformatics*

# HELIQUEST: a web server to screen sequences with specific α-helical properties

Romain Gautier*, Dominique Douguet, Bruno Antonny and Guillaume Drin

Institut de Pharmacologie Moléculaire et Cellulaire, Université de Nice Sophia Antipolis and CNRS, 660 route des Lucioles, 06560 Valbonne, France

Received on April 17, 2008; revised on July 16, 2008; accepted on July 25, 2008

Advance Access publication July 28, 2008

Associate Editor: Anna Tramontano

**ABSTRACT**

**Summary:** HELIQUEST calculates the physicochemical properties and amino acid composition of an α-helix and screens databank to identify protein segments possessing similar features. This server is also dedicated to mutating helices manually or automatically by genetic algorithm to design analogues of defined features.

**Availability:** http://heliquest.ipmc.cnrs.fr

**Contact:** gautier@ipmc.cnrs.fr

## 1 INTRODUCTION

Several proteins bind to membranes via a small amphipathic helix with one face made of hydrophobic residues that insert between the lipid acyl chains and the other one containing polar residues that interact with the lipid polar heads and solvent (Cornell and Taneva, 2006). This structural motif and its mode of membrane binding could appear simplistic, compared to well-structured domains (PH, PX, FYVE) that recognize specific lipids (Lemmon, 2008). However, depending on their sequence, amphipathic helices have different properties: some form holes in membrane as certain antimicrobial peptides (Dathe and Wieprecht, 1999); others permit proteins of vesicular transport to modify or detect membrane shape (McMahon and Gallop, 2005). For instance, ArfGAP1 uses two amphipathic ALPS helices (*A*mphipathic *L*ipid *P*acking *S*ensor) to recognize the curved membrane of nascent transport vesicles molded by the COPI coat and triggers in turn the depolymerization of this coat (Mesmin *et al*., 2007). The ALPS helix binds specifically to curved membrane, because its polar face is rich in serine and threonine and contains almost no basic residues. Unable to interact electrostatically with lipids, this helix binds to membrane exclusively through the insertion of its hydrophobic face between lipid acyl chains. This insertion is favored when the outer membrane leaflet is expanded by positive curvature. Introducing two basic residues in the polar face was enough to create a curvature-insensitive helix interacting with flat membranes (Drin *et al*., 2007). This result illustrated how critical are the composition and physicochemistry of an amphipathic helix for its function and helped us to develop an algorithm to find curvature-sensitive proteins.

We failed to find ALPS-containing proteins by sequence similarity (using BLAST or related algorithm) other than ArfGAP1 orthologues. Thus, we developed an algorithm that extracts from the SWISSPROT database protein segments whose sequence, although different from ALPS, exhibits once considered as helical, similar composition and physicochemical properties. Our consideration was that numerous amino acid combinations in a sequence could result in an amphipathic helix with a S/T-rich polar face. Eventually, we demonstrated that three proteins, selected among ∼400 hits, use an identified ALPS-like segment, folded into an α-helix, to recognize curved membranes (Drin *et al*., 2007). This was validated by others studies (Alber *et al*., 2007; Drin *et al*., 2008).

Tools such as MPEX (http://blanco.biomol.uci.edu/mpex/) or Amphipaseek (Sapay *et al*., 2006), devoid of screening capacity, are powerful to localize amphipathic segments but only in proteins known by the user. We created HELIQUEST, based on our algorithm, to allow users to determine online the features of known helices (amphipathic, transmembrane, etc.) and use the results as a starting point to extract putative equivalent helices in unexpected proteins. An additional module was developed to manually mutate a helix or to automatically design analogues by genetic algorithm.

## 2 METHODS

### 2.1 Sequence analysis and database screening

A sequence submitted by the user, considered as helical, is analyzed by a sliding window (14–54 aa, i.e. up to three repeats of a complete helical wheel of 18 aa). The analysis module displays for each generated segment a table reporting its net charge $z$ (at pH = 7.4), mean hydrophobicity <H> and hydrophobic moment <μH> (Eisenberg *et al*., 1982) calculated with a standard hydrophobicity scale (Fauchere and Pliska, 1983), as well as statistics on its composition (percentage or enumeration of specific residues). A helical wheel representation of each segment with its <μH> vector is downloadable.

If user analyzes sequences with an 18 aa window, each table displays a link to the screening module. Values calculated on the selected segment and displayed in a new web page help the user to define an interval of <H>, <μH> and $z$ values and to specify if the protein segment to be extracted must contain a minimal number of polar residues (E, D, K, R, S, T, N, H, Q + G), of charged residues (E, D, K, R) and of specific polar residues (S, T, N, Q, H). Protein segments containing proline at their ends, cysteine or both can be either accepted or excluded. The screening module is directly accessible if the parameters are known. For example, we identify ALPS-like motifs by limiting <H> between 0.28 and 0.6 and $z$ between −1 and 2 with <μH> superior to 0.35, a maximum of four charged residues and a sum of serine, threonine and glycine superior to 6; cysteine was excluded.

User screens either personal or annoted SWISSPROT databases (Boeckmann *et al*., 2003); in this case, one can discard poorly

*To whom correspondence should be addressed.

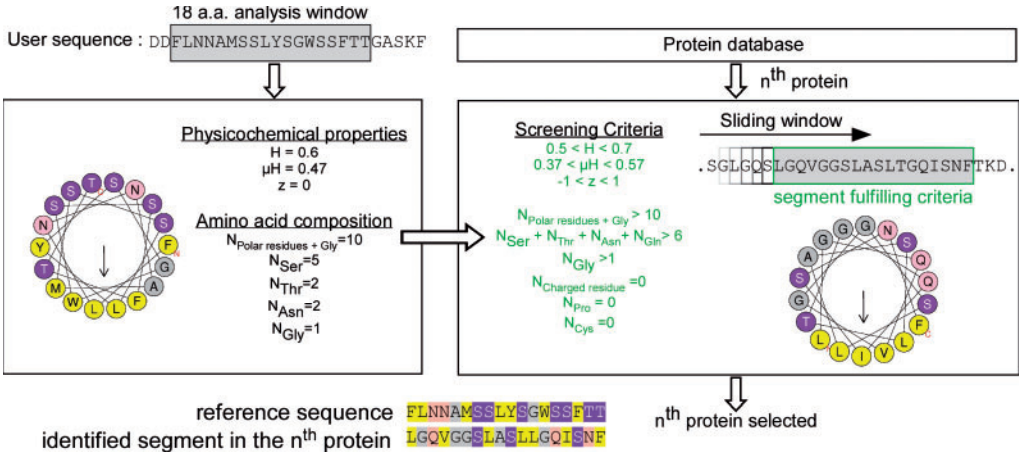


**Fig. 1.** Simplified description of the screening strategy. Sequence analysis provides $\alpha$-helical-related physicochemical parameters and statistics on amino acid composition that are used as reference criteria to screen protein databases. Note that this procedure identifies sequences that, one considered as helical, have features similar to those of reference sequence despite a completely different primary sequence.

defined proteins. The algorithm slides an 18 aa window along each protein sequence of the databank and save segments fulfilling the required criteria (Fig. 1). A procedure exists to refine the screening of amphipathic helices by discarding sequences that despite a high $<\mu H>$ value do not have well-defined polar and nonpolar faces. This algorithm detects the existence of an uninterrupted hydrophobic face of at least five residues adjacent on a helical wheel. If such a face exists, it verifies whether the facing residues are polar.

An output text file lists the proteins containing at least one positive segment, whose sequence, localization, physicochemical features and content in amino acid are reported. If several segments in one protein overlap or are adjacent, they are merged into a unique sequence, reported in a second file. Corresponding PDF files contain helical wheel representations of all sequences. A decision tree, integrating results from TMHMM (Krogh *et al*., 2001), PSIPRED (Jones, 1999) and a discriminant analysis performed on lipid-binding helices, order sequences in six classes. A sequence is classified for example as a helix, a lipid-binding helix or as a non-relevant sequence with a high propensity to form a $\beta$-sheet. With appropriate screening parameters, our procedure was found to identify transmembrane segments from the MPtopo database (Jayasinghe *et al*., 2001) and highly amphipathic helices (with a $<\mu H>$ superior to 0.6) from a subset of non-redundant PDB with a positive predictive value of 95 and 86%, respectively. Finally, we noted that the screening procedure was able to identify a majority of lipid-binding helices extracted from a small dataset of known perimembrane proteins.

## 2.2 Helix mutation

Any characterized helix can be mutated either manually (and reanalyzed to examine how the mutation changes its features) or automatically by genetic algorithm (GA). As $<H>$, $<\mu H>$ and $z$ are interdependent properties, changing manually one property without modifying others is difficult (Dathe and Wieprecht, 1999). The GA-based module allows modifying independently or simultaneously these parameters with a minimal number of mutations. Alternatively, user can impose an amino acid composition within an $\alpha$-helix under defined $<H>$ and $<\mu H>$ constraints. Generating sequences *de novo* with precise features is possible.

## 2.3 Implementation

HELIQUEST, written in Python 2.5, is organized as interconnected CGI programs and use R language to draw helical wheel.

*Funding*: Agence Nationale pour la Recherche; the CNRS (RISC).

*Conflict of Interest*: none declared.

## REFERENCES

Alber,F. *et al*. (2007) Determining the architectures of macromolecular assemblies. *Nature*, **450**, 683–694.

Boeckmann,B. *et al*. (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res*., **31**, 365–370.

Cornell,R.B. and Taneva,S.G. (2006) Amphipathic helices as mediators of the membrane interaction of amphitropic proteins, and as modulators of bilayer physical properties. *Curr. Protein Pept. Sci.*, **7**, 539–552.

Dathe,M. and Wieprecht,T. (1999) Structural features of helical antimicrobial peptides: their potential to modulate activity on model membranes and biological cells. *Biochim. Biophys. Acta.*, **1462**, 71–87.

Drin,G. *et al*. (2007) A general amphipathic alpha-helical motif for sensing membrane curvature. *Nat. Struct. Mol. Biol.*, **14**, 138–146.

Drin,G. *et al*. (2008) Asymmetric tethering of flat and curved lipid membranes by a golgin. *Science*, **320**, 670–673.

Eisenberg,D. *et al*. (1982) The helical hydrophobic moment: a measure of the amphiphilicity of a helix. *Nature*, **299**, 371–374.

Fauchere,J. and Pliska,V. (1983) Hydrophobic parameters {pi} of amino-acid side chains from the partitioning of *N*-acetyl-amino-acid amides. *Eur. J. Med. Chem.*, **8**, 369–375.

Jayasinghe,S. *et al*. (2001) MPtopo: a database of membrane protein topology. *Protein Sci.*, **10**, 455–458.

Jones,D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.*, **292**, 195–202.

Krogh,A. *et al*. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.*, **305**, 567–580.

Lemmon,M.A. (2008) Membrane recognition by phospholipid-binding domains. *Nat. Rev. Mol. Cell Biol.*, **9**, 99–111.

McMahon,H.T. and Gallop,J.L. (2005) Membrane curvature and mechanisms of dynamic cell membrane remodelling. *Nature*, **438**, 590–596.

Mesmin,B. *et al*. (2007) Two lipid-packing sensor motifs contribute to the sensitivity of ArfGAP1 to membrane curvature. *Biochemistry*, **46**, 1779–1790.

Sapay,N. *et al*. (2006) Prediction of amphipathic in-plane membrane anchors in monotopic proteins using a SVM classifier. *BMC Bioinformatics*, **7**, 255.