Supplementary Figures for

The TMCrys server for supporting crystallization of transmembrane proteins

Julia K. Varga and Gábor E. Tusnády*

"Momentum" Membrane Protein Bioinformatics Research Group, Institute of Enzymology, Research Center of Natural Sciences, Hungarian Academy of Sciences, H-1117 Budapest, Magyar tudósok körútja 2.

Supplementary Figure 1	2
Supplementary Figure 2.	3

Supplementary Figure 1.



Overview of the server. The submitted sequence is stored in a MYSQL database. The HPC (high performance computing) cluster and the frontend server communicates via SOAP requests by an Apache AXIS2 server. Once the sequence arrives on the cluster, features are calculated by running several programs and scripts, then fed into an R script that makes the XGBoost prediction. The process is monitored by the AXIS2 server and once finished, the results are sent back to the front-end server, stored in the MYSQL database and displayed.

Supplementary Figure 2.

Α	-	5132_D				-	В	*	PLPR2_HUMAN				-
	Prediction	Prediction Purification Purification O.976 O.983 O.851 O.937 Reliability of prediction: O.894 Sequence GAGVAFGSFDDSF5LASLRAVLAEFISTLLEVFA0VGSALAYAKLTSDAALDTPGLVALAVCHGFALFVAVAIGMISGGHVNPAVTFGLAVGG0ITVITGVFYvIJAQLLGSTAACFLLKVVTGGL AVPTHSVAAGLESIEGVWEITITFALVYTVATADPKKGSLGTIAPLATGLIVGANILAAGPFSGGSMIPARSFGPAVAAGDFSGHWNPAVTFGLAVGG0ITVITGVFYWIJAQLLGSTAACFLLKVVTGGL						Prediction	Solubilization Purification Crystallization Whole process Reliability of prediction:	0.884	0.789 0.679 0.218 0.562		
	Sequence							Sequence	quence MAGGRPHLKRSFSIIPCFVFVESVLLGTVILLAYRLEFTDTFPVHT0GFCCYDSTYAKPY6GEAASRVPPALYYALVTAGPTLTILLGELARAFFPAPPSAVPVIGESTIVSGACCRFSPP VRRLVRFLGVYSFGLFTTTIFANAG0VTGNPTPHFLSVCRPHYTALGCLPPSPDRPGPDRFVTDGGACAGSPSLVAARRAFPCKDAALCAYAVTYTAWYVTLVFRVKGSRLVKPSLCLAL LCPAFL0VVSFGLFTTIFANAG0VTGNPTPHFLSVCRPHYTALGCLPPSPDRPGPDRFVTDGGACAGSPSLVAARRAFPCKDAALCAYAVTYTAWYVTLVFRVKGSRLVKPSLCLAL LCPAFL0VVSFGLFTTIFANAG0VTGNPTPHFLSVCRPHYTALGCLPPSPDRPGPDRFVTDGGACAGSPSLVAARRAFPCKDAALCAYAVTYTAWYVTLVFRVKGSRLVKPSLCLAL LCPAFL0VVSFGLFTTIFANAG0VTGNPTPHFLSVCRPHYTALGCLPPSPDRPGDRLSPWEDLGQAPTNDSPLEKNPSAGRIRHRHGSPHPSRTAPAVAT pology 111111111111111111111111111111111111				
	Topology	logy IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII						Topology					
	Similar prot	Imilar proteins in TSTMP											
		30	TSTMP_004265	151MP_000977	151MP_001597	TSTMP_002784		Similar prote	ans in TSTMP	No similar 2D optor found in	TCTMD		
		Modelable	No similar modelable entry	found in TSTMP.					Modelable	No similar modelable entry	found in TSTMP.		
	Similar prot	eins in TargetTrack	No similar target entry foun	d III I STMP.					Target	TSTMP_000015	TSTMP_004232	TSTMP_000439	TSTMP_000890
		0	CESG-GO.22702 CESG-GO.16964 CSMP-10705 RSGI-hso002001025.1 TEMIMPS-G0.20548	CESG-GO.20754 CESG-GO.20753 CSMP-42785 CSMP-2002 NYCOMPS-GO.11653	NYCOMPS-G0.11649 NYCOMPS-G0.11666 NYCOMPS-G0.11643 NYCOMPS-G0.11662 NYCOMPS-G0.11586	NYCOMPS-GO.11583 NYCOMPS-GO.11625 NYCOMPS-GO.11626 SECSG-C01G6.1A NYCOMPS-GO.11500		Similar prote	eins in TargetTrack	SECSG-T06D8.3	SECSG-F53C3.13	151MP_002161	
		R5GI-mm010002852.1 NYCOMP5-G0.11527 TEMIMF5-G0.15032 CSMP-11715 TEMIMF5-G0.15032 CSMP-11715 TEMIMF5-G0.15040 NYCOMP5-G0.1664 CESG-G0.9729 CSMP-4063 CESG-G0.9730 CHTSB-YPF192W TEMIMF5-G0.15289 CESG-G0.9808 SESC5-G32C4.2 CSMP-1170.5 TEMIMF5-G0.15288 NYCOMF5-G0.1624	NYCOMPS-GO.11597 NYCOMPS-GO.11628 CSMP-1171S NYCOMPS-GO.11664 CSMP-4063	7 NYCOMPS-G0.11627 NYC 8 NYCOMPS-G0.11650 LJCS NYCOMPS-G0.11650 LJCS NYCOMPS-G0.11651 CSN SGPP-tmaj005431AAA CSN NYCOMPS-G0.11616 CES NYCOMPS-G0.11617 TMI 7 NYCOMPS-G0.11642 CSS	NYCOMPS-G0.11655 LCS-283289 TMPC-G0.103988 NYCOMPS-G0.11606 CSMP-24645 CSMP-24645 CSMP-24645 CSMP-2625 CESG-G0.103987 CSMP-24545 CESG-G0.103987		Features	Length		343			
			CHTSB-YPR192W CESG-GO.9808 CSMP-11705					GRAVY		4.507			
			NYCOMPS-GO.11667 NYCOMPS-GO.11624					Isoelectric point		9.64			
			CESG-GO.6573 SECSG-F40F9.9 CESG-GO.10496	NYCOMP5-G0.11602 NYCOMP5-G0.11656 NYCOMP5-G0.11608	NYCOMPS-G0.11648 NYCOMPS-G0.11636 NYCOMPS-G0.11665	NYCOMPS-GO.11661 NYCOMPS-GO.11589			OB score		-4.805		
			CESG-GO.8072 CESG-GO.16539	NYCOMPS-GO.11609 NYCOMPS-GO.11657	NYCOMPS-GO.11659 NYCOMPS-GO.11637				Instability index		53.07		
	Features								Half life		30		
		Length		238					Ratio of buried residues	5	1.287		
		Log(molecular weight) 4.375							Average solvent accessi	ble surface area	0.292		
		GRAVY		1.013			С						
		Isoelectric point		5.67			Ŭ	notTMP	PRKN_HUMAN				-
		OB score		-1.071				Sequence					
		Instability index 25.26				MIV	MIVFVRFNSSHGFPVEVDSDTSIFQLKEVVAKRQGVPADQLRVIFAGKELRNWTVQNCDLDQQSIVHIVQRPWRKGQEMNATGGDDRNAAGGCEREPQSITRVDLSSSVLPGDSVGLAVILHTDSR						
		Ratio of buried residues		2.778				HVICLDCFHLYCVTRLNDRQFVHDPQLGYSLPCVAGCPNSLIKELHHFRILGEQYNRYQQYGAEECVLQMGGVLCPRPGCGAGLLPEPDQRKVTCEGGNGLGCGFAFCRECKEAYI TTTDAYRVDFRAAFGARWFAASKFTIKKTIKPCPRCHVPVFKNGGCMHMKCPOPDCRLFWCMNCGCFWNRYCMGDHWFDV					GFAFCRECKEAYHEGECSAVFEASG
		Average solvent accessit	ble surface area	0.233				. Tryattyp	and a suppression of the state				

Main types of output of TMCrys server. (A) Green box marks the proteins that are predicted to be transmembrane by CCTOP and predicted to be crystallizable by TMCrys. (B) Rex box marks the proteins that are predicted to be transmembrane by CCTOP and predicted to be non-crystallizable by TMCrys. (C) Yellow box marks the proteins that are not predicted to be transmembrane by CCTOP. On these proteins, no further calculations are made. *These predictions can be accessed at http://tmcrys.enzim.ttk.mta.hu/result/examplejob.*