

# BBF RFC 96: The AutoAnnotator - Standardized annotation of protein-encoding BioBricks

Christopher Wolf, Florian Albrecht, Johanna Brüggenthies, Andreas Brunner, Rosario Ciccone, Katrin Fischer, Fabian Froehlich, Louise Funke, Volker Morath, Ingmar Polte, Leonie Reichart, Philipp Schneider, Jeffery Truong

September 25, 2013

## 1. Purpose

For protein-encoding BioBricks a large number of useful parameters, such as the amino acid sequence or the molecular weight, can easily be computed. Furthermore alignments against databases and predictions of various features, for instance transmembrane regions, are very interesting and informative when viewing a part. However, this information is not presented or very hard to find in most part descriptions and requires the user to use external websites to obtain it. To resolve this issue we propose a standardized table containing this information. For this purpose we developed a web tool, called the AutoAnnotator, in JavaScript gathering this information and providing the standardized table for the user to copy.

## 2. Relation to other BBF RFCs

This RFC extends BBF RFC 89: The minimum information for a qualified BioBrick.

## 3. Copyright Notice

Copyright (C) The BioBricks Foundation (2013). All Rights Reserved.

## 4. Use of the AutoAnnotator

The AutoAnnotator can be found at: <http://2013.igem.org/Team:TU-Munich/Results/AutoAnnotator>

When adding a new protein-encoding BioBrick to the Registry, the user **MUST** apply the AutoAnnotator to the part and add the resulting table to the description of his/her part. The user **MUST** check, whether the computations were successful and whether the results (in particular the automatically determined Open Reading Frame) are as expected. For proteins not present in the data base used for the predictions, the table will initially be displayed without these, until they have been calculated, which should take no more than two days. If this is the case, the user **SHALL** reapply the AutoAnnotator to his/her part after two days and add the updated table to his/her part description.

If the user has additional information on the categories in the table, e.g. has experimental evidence of a transmembrane region, he SHOULD add this to the table appropriately, designating the nature of the evidence. The format of the table SHOULD be as below.

Protein data table for BioBrick Bba_K801060 automatically created by the BioBrick-AutoAnnotator version 1.0									
<b>Nucleotide sequence in RFC 10:</b> (underlined part encodes the protein) GTCACACA <u>TGCGTCGT</u> ... <u>TCGAAAAATAA</u>									
<b>ORF</b> from nucleotide position 8 to 1705 (excluding stop-codon)									
<b>Amino acid sequence:</b> (RFC 25 scars in shown in bold, other sequence features underlined; both given below) 1 MRRSANYQPSIWDHDFLQSLNSNYTDEAYKRRAEELRGKVKIAIKDVEIPLDQLDLIDNLQRLGLAHRFETEIRNINLNNIYNNNKDYNWRKENLYATSLE 101 FRLLRQHGYPVSEQEVFNGFKDDQGGFICDDFKGILSLHEASYSLEGESIMEEAWQFTSKHLKEVMISKNMEEDVFVAEQAKRALEPLLHWKVPMLERARW 201 FIIHYERREDKNHLLLELAKMEFNTLQAIYQEELKEISGWKDTGLGEKLSFARNRLVASFLWSMGIAFEPQFAYCRRVLTISIALITVIDDIYDYVYGT 301 DELEIFTD AVERWDIN YALKHLPGYMRMCF LALYNFVNEFAYVYVVKQQDFD LLSIKNAWLGLIQAYLVEAKWYHSKYTPKLEEYLENGLV SITGPLIIT 401 ISYLSGTNPIIKKELEFLESNPDIVHWSSKIFRLQDDLGTSSEDEIQRGDVPKSIQCYMHETGASEEVARQHIKDMMRQMWKKNAYTADKDSPLTGTTE 501 FLLNLVRMSHFMYLHGDGHGVQNETIDVGFTLLFPQIPILEDKHMFASTASPGTKGAWSHHPQFEK*									
<b>Sequence features:</b> (with their position in the amino acid sequence, see the <a href="#">list of supported features</a> ) RFC25 scar (shown in bold): 556 to 557 Strep-tag II: 559 to 566									
<b>Amino acid composition:</b>									
Ala (A)	33 (5.8%)	Cys (C)	4 (0.7%)	His (H)	18 (3.2%)	Met (M)	16 (2.8%)	Thr (T)	26 (4.6%)
Arg (R)	25 (4.4%)	Gln (Q)	24 (4.2%)	Ile (I)	39 (6.9%)	Phe (F)	28 (4.9%)	Trp (W)	14 (2.5%)
Asn (N)	27 (4.8%)	Glu (E)	48 (8.5%)	Leu (L)	64 (11.3%)	Pro (P)	17 (3.0%)	Tyr (Y)	27 (4.8%)
Asp (D)	34 (6.0%)	Gly (G)	29 (5.1%)	Lys (K)	36 (6.4%)	Ser (S)	33 (5.8%)	Val (V)	24 (4.2%)
<b>Amino acid counting</b>				<b>Biochemical parameters</b>					
Total number:		566		Atomic composition:		C <sub>3002</sub> H <sub>4586</sub> N <sub>778</sub> O <sub>868</sub> S <sub>20</sub>			
Positively charged (Arg+Lys):		61 (10.8%)		Molecular mass [Da]:		66105.3			
Negatively charged (Asp+Glu):		82 (14.5%)		Theoretical pI:		5.38			
Aromatic (Phe+His+Try+Tyr):		87 (15.4%)		Extinction coefficient at 280 nm [M <sup>-1</sup> cm <sup>-1</sup> ]:		117230 / 117480 (all Cys red/ox)			
<b>Plot for hydrophobicity, charge, predicted secondary structure, solvent accessibility, transmembrane helices and disulfid bridges</b> <span style="float:right">Hide</span>									
<input checked="" type="checkbox"/> Moving average over 5 amino acids for hydrophobicity (H) <input checked="" type="checkbox"/> Moving average over 5 amino acids for charge (C) <input checked="" type="checkbox"/> Predicted disulfid bridges (S) with the number of the bridge in the center <input checked="" type="checkbox"/> Predicted transmembrane helices (H) <input checked="" type="checkbox"/> Predicted secondary structure: Helices (H) and beta-strands (B) <input checked="" type="checkbox"/> Predicted solvent accessibility: Exposed (E) and buried (B) residues									
<b>Codon usage</b>									
Organism:	<i>E. coli</i>	<i>B. subtilis</i>	<i>S. cerevisiae</i>	<i>A. thaliana</i>	<i>P. patens</i>	Mammals			
Codon quality (CAI):	good (0.71)	good (0.75)	good (0.69)	good (0.78)	excellent (0.80)	good (0.68)			
<b>Alignments</b> (obtained from <a href="#">PredictProtein.org</a> )									
SwissProt:	Q8L5K3 (99% identity on 554 AAs), Q93X23 (54% identity on 536 AAs)								
TrEMBL:	A7BG59 (99% identity on 554 AAs), Q8L5K1 (96% identity on 554 AAs)								
PDB:	1nlz (44% identity on 530 AAs), 1n20 (44% identity on 525 AAs)								
<b>Predictions</b> (obtained from <a href="#">PredictProtein.org</a> )									
<b>Subcellular Localization</b> (reliability in brackets)				<b>Gene Ontology</b> (reliability in brackets)					
Archaea: cytosol (100%)		Bacteria: cytosol (95%)		Eukarya: nucleus (15%)		Molecular Function Ontology: GO:005051 (24%), GO:0034768 (24%)			
						Biological Process Ontology: GO:0071485 (19%), GO:0009611 (18%)			
<b>Predicted features:</b>									
Disulfid bridges:		-							
Transmembrane helices:		-							
The BioBrick-AutoAnnotator was created by <a href="#">TU-Munich 2013</a> iGEM team. For more information please see the <a href="#">documentation</a> . If you have any questions, comments or suggestions, please leave us a <a href="#">comment</a> .									

When viewing a table created by the AutoAnnotator the user SHOULD check, that the used open reading frame is correct. Additionally, since there might be a newer version of the AutoAnnotator and the quality of the predictions will improve as more structures are analyzed, the user MAY want to re-run the AutoAnnotator.

## 5. Authors' Contact Information

Christopher Wolf, [christopher.guy.wolf@web.de](mailto:christopher.guy.wolf@web.de)  
Florian Albrecht, [albrecht.flo@gmx.net](mailto:albrecht.flo@gmx.net)  
Johanna Brüggenthies, [johanna.brueggenthies@kabelmail.de](mailto:johanna.brueggenthies@kabelmail.de)  
Andreas Brunner, [brunner.andreasdavid@web.de](mailto:brunner.andreasdavid@web.de)  
Rosario Ciccone, [r.ciccone@gmx.de](mailto:r.ciccone@gmx.de)  
Katrin Fischer, [katrin.fischer@email.de](mailto:katrin.fischer@email.de)  
Fabian Froehlich, [fabian.froehlich@mytum.de](mailto:fabian.froehlich@mytum.de)  
Louise Funke, [louise.funke@web.de](mailto:louise.funke@web.de)  
Volker Morath, [v.morath@tum.de](mailto:v.morath@tum.de)  
Ingmar Polte, [ingmar.polte@mytum.de](mailto:ingmar.polte@mytum.de)  
Leonie Reichart, [leonie.reichart@gmail.com](mailto:leonie.reichart@gmail.com)  
Philipp Schneider, [philipp.schneider@hotmail.de](mailto:philipp.schneider@hotmail.de)  
Jeffery Truong, [jeffery\\_truong@mytum.de](mailto:jeffery_truong@mytum.de)

## References

Mubing Zhou, Junqiu Zhang and Jiankui He “BBF RFC 89: The minimum information for a qualified BioBrick“

Full documentation of the AutoAnnotator at <http://2013.igem.org/Team:TU-Munich/Results/Software>