J Mol Med (1998) 76: 77–79

Bioinformatics: Bits and Bytes

by Peer Bork

Characterization of targeting domains by sequence analysis: glycogen-binding domains in protein phosphatases

Peer Bork · Thomas Dandekar · Frank Eisenhaber · Martijn Huynen

The prediction of functional features from sequence data can powerfully complement experimental analyses. Domains are delineated in a first step. Subsequent analysis can provide functional information that even may include precise structural details. This contribution describes the way in which sequence analysis has used multiple alignment, blast and reverse blast searches, and other, more sophisticated techniques to produce a detailed picture of the glycogen-targeting domain in the glycogen-binding subunit of the protein phosphatase 1 (PP1) family that includes its three-dimensional fold.

Abbreviations *PP* Protein phosphatase · *PTG* Protein targeting to glycogen

P. Bork (✉) · T. Dandekar
F. Eisenhaber · M. Huynen
EMBL, Meyerhofstrasse 1,
D-69012 Heidelberg, Germany

Please send articles to:
Peer Bork
Max-Delbrück-Center
for Molecular Medicine (MDC)
Robert-Rössle-Strasse 10
D-13122 Berlin, Germany
and:
EMBL
Meyerhofstrasse 1
D-69117 Heidelberg, Germany
E-mail: bork@embl-heidelberg.de
http://www.embl-heidelberg.de/~bork/

Printen et al. [1] identified a novel glycogen-sensitive subunit of protein phosphatase 1 (PP1G) which they called protein targeting to glycogen (PTG) because it mediates the association of certain proteins with glycogen. PTG is in fact specific for the primary enzymes involved in hormonal regulation of glycogen metabolism in insulin-sensitive tissues [1]. Sequence analysis of PTG along with other PP1Gs have allowed us (a) to predict the location and three-dimensional structure of the putative high-affinity glycogen-binding sites (Fig. 1) and (b) to localize binding sites for the catalytic subunit (PP1C) of PP1.

From the entire yeast genome four PP1G homologues have been experimentally implicated in glycogen synthesis [2], each of which manifests only two conserved regions in common with the animal PP1Gs. We subjected one of these regions, which is located in the central part of PTG, to iterative motif and profile searches [3]. In addition to human PTG, a distinct human homologue, EST R10183, was retrieved. This is most probably a human paralogue of PTG, meaning that the protein is not functionally identical to PTG, inasmuch as the homologous glycogen-binding domain would act in concert with other domains to effect a function different from that of PTG. Our newly identified paralogue is, incidentally, distinct from the novel protein phosphatase that has recently been sequenced [13].

Furthermore, we have identified numerous high-affinity starch-binding domains of distinct glycohydrolases (Fig. 1). The N- and C-terminal location of the domains in some of the glycohydrolases, along with the three-dimensional structures that have been solved in several instances, allowed us to define the domain borders, such that a β-sandwich fold can be predicted on the basis of alignments and secondary structure predictors such as the PhD program [14]. Strong independent evidence supporting this prediction is provided by the identification of the domain in a subset of glycohydrolases in which the exact three-dimensional structure has been experimentally determined (deposited in the Brookhaven files 1KUL, a glucoamylase, and 1CGT, a glycosyltransferase). Mapping of the conserved residues on the predicted structure furthermore reveals the exposed polar amino acids that would be expected to interact with glycogen (Fig. 1). A detailed analysis of the domain including color figures of the structure are given at the web site http://www.bork.embl-heidelberg.de/pp1g/pp1g.html.

Both the recognition of the homologues with known three-dimensional structure and computer retrieval of the human paralogue required accurate determination of the domain boundaries and also depended on iterative procedures which assign functionality to the relevant peptide sequence. Without such accurate procedures the glycogen targeting ability of PP1G is not ascribed to a particular domain of the complete protein [1] or incorrectly ascribed to either a stretch of 13 amino acid residues [11] or to multiple amino acid stretches at the N-terminus characterized by conserved hydrophobic motifs [12]. Although the direct alignment of five sequences, including a novel cloned PP1-binding protein, did reveal

```
2D prediction:              aaaaaaaaaaaa   bbbbbbb                  ddddddd           eeeeeeeee        ffff       ggg
PTG/Mouse    U89924  144 RTVTGTVKVKNVSFE-KKVQVRIT-(0)-FDTWK-(18)-SDTFSFAIDLPR-( 4)-EEKIEFCISYHAN- (1)-RIFWDNNEGQNYRI
PP1M/Human   X78578  140 TSIKGIIRVLNVSFE-KLVYVRMS-(0)-LDDWQ-(17)-TDQFSFKIVLVP- ( 5)-GSKVEFCIRYETS- (1)-GTFWSNNNGTNYTF
PP1L/Rat     S80360  142 KAIAGTVKVQNLAFE-KVVKIRMT-(0)-FDTWK-(18)-RDTFSFDISLPE- ( 4)-YERMEFAVCYECN- (1)-QSYWDSWKGKNYRI
PIG2_Yeast   P40187 406 MVIIGKVFVKNIYFD-KKIIVRYT-(0)-WDAWR-(21)-MDIFKFSIDDIH- ( 8)-ISQLEFCIQYLTW- (6)-KEYWDNWDSANYKI
GIP2_Yeast   P40036 437 RLIAGRILVKNIFYD-KRVVVRYT-(0)-WDSWR-(21)-MDIFHFIIDDVS- ( 4)-RGKLEFCIHYSTR- (5)-EEYWDNWNGNNYKV
GAC1_Yeast   P28006 257 SKITGLVYVKNLSFE-KYLEIKFT-(0)-FNSWR-(16)-VDEFKFTIDLNS- (20)-PLNIELCCRYDVN- (1)-ETYYDNWNGKNYHL
PIG1_Yeast   Q06216 222 CQIWGLIFVNNLNFE-KKIEIKFT-(0)-LNNWA-(16)-VDEFKFIIDISA- (22)-LLNLQFCCRYDVN- (4)-RSFYDNWDYKNYEI
AMYG_Rhior   P07683  44 STFSGKIYVKNIAYS-KKVTVIYA-(2)-SDNWN-(18)-YEYWTFSASING- ( 0)---IKEFYIKYEVS- (1)-KTYYDNWNSANYQV
AMYA/Bacuc   X53373 976 EGNLVTIYYKK-GFD--TPYMHYR-(2)-GGEWT-(11)-IAGY-SKLTVDI- ( 0)-REASKLEVAFNNG- (0)-RGAWDSDQENNYLF
AMYC/Strli1  Z86113  20 PGNTATVFYYTKTKNWDRYNLHYA-(2)-GGSWT-(11)-CTDW-VKRTVPL- ( 0)-GSAEGLRATFNNG- (0)-SGTWDNWGGENYAL
AMYC/Strli2  Z86113 128 DGNRATVYYSTRTLGWTTANIHYR-(2)-GGSWT-(11)-CAGW-WKKDVDL- ( 0)-GAATSLTAAFNNG- (0)-NGVWDNWKGADYTL
GLYC/Bacci   1cgt   583 DQVTVRFVVNNASTTLGQNLYLTG-(4)-LGNWS-(16)-YPTWYYDVSVPA- ( 0)-GKQLEFKFFKKNG- (1)-TITWESGSNHTFTT
AMYA/Aspni   1kul     5 TAVAVTFDLTA-TTTYGENIYLVG-(4)-LGDWE-(15)-DPLWYVTVTLPA- ( 0)-GESFEYKFIRIES- (2)-SVEWESDPNREYTV
AMYA/Bacsp   U22045 514 KTSNVTFTVNNATTTSGQNVYVVA-(4)-LGNWN-(11)-YPTWKATIALPQ- ( 0)-GKAIEFKFIKKDQ- (2)-NVIWESTSNRTYTV
AMYB_Bacce   P36924 447 VPLLSTIVVKNLPTTIGDTVYITG-(4)-LGSWT-(12)-SNDWRGNVVLPA- ( 0)-ERNIEFKAFIKSK- (3)-VKSWQTIQQSWNPV
2D in 1KUL:                 aaaaaaaaaaa       bbbbbb                dddddddddddd       eeeeeeeeee          fff    ggggg
```

Fig. 1 Alignment of PTG and some related PP1G subunits with selected starch-binding domains of glycohydrolases. *First column*, protein and species abbreviations taken from or adapted to the SWISS-PROT database; *second*, *third columns*, database accession numbers and positions are given in the respective sequences for unique identification; *numbers between the blocks*, the intervening residues. The starch-binding domains [6, 7] of glucoamylase G2 from *Aspergillus nidulans* (Protein Data Bank code 1KUL) and cyclodextrin-glycosyltransferase from *Bacillus circulans* (Protein Data Bank code 1CGT) are taken from the Protein Data Bank of known three-dimensional structures. Secondary structure prediction of the animal PP1Gs (*above alignment*) is in agreement with the β-strands a–g obtained from 1KUL (*beneath alignment*). *Bold*, hydrophobic positions. Several of the conserved residues (*highlighted*) cluster together in the structure with the polar side chains being candidates for direct interaction with glycogen. The similarity of PP1Gs and the starch-binding domain of Amyg_Rhior can be elaborated by standard methods such as Blast [2, 8] yielding a probability of a chance match P=4.6.10^{-8}. A similarity of the N-terminal domain in Amyg_Rhior and other starch-binding domains of glycohydrolases has already been noted [9]. Motif and profile searches (for strategies see [3]) quantify these findings and identify a large super family of β-sandwich domains resembling structurally related fibronectin type 3 repeats that also occur as carbohydrate-binding domains in various glycohydrolases (see [10] and references therein). For more data including the detailed search protocol a full-length alignment with more family members and three-dimensional models of the domains see http://www.bork.embl-heidelberg.de/pp1g/pp1g.html

a putative glycogen binding domain (residues 128–229 according to the numbering of human PP1M) [13], the more elaborate sequence analysis detailed here is indispensable for statistical verification of this finding (*P* values and MACAW scores) and determination of boundaries and three-dimensional structure (Fig. 1 and below).

The second region of significant similarity, which is shared by all PP1Gs and has the consensus sequence R/K-R/S/N-V-x-F-A/D (where x is any amino acid and a slash indicates alternative residues for the given position), starts at position 60 in PTG (probability of chance matching P=1.7×10^{-14} as measured by the MACAW program [4]). Significantly, the binding site for PP1C on the rabbit PP1 Gm subunit has been located in a region that includes this motif [5]. During preparation of this manuscript, moreover, Egloff et al. [15] confirmed this second prediction by reporting direct structural analyses of the interaction of the peptide motif from PP1Gm with rabbit skeletal muscle PP1C; the functional disruption of the interaction was furthermore established upon mutation within the motif.

Our sequence analysis suggests that at the molecular level common mechanisms exist for all PP1Gs whereby they direct the associations of appropriate enzymes with both glycogen and PP1-C. Other features, such as phosphorylation sites and tissue- and compartment-specific activities of distinct PP1Gs [1], are not shared by all family members and are therefore expected to be specified by nonconserved sequences elsewhere in the molecules.

An ongoing debate [15] centers on whether regulatory subunits modulate substrate specificity of PP1C by altering PP1C conformation or by the specific targeting of substrates. At least with regard to PP1G, we now present strong evidence, based on sequence analysis, that substrate specificity is achieved through targeting, as we can relate the sequence motif to known crystal structures of domains directly involved in glycogen binding.

References

1. Printen JA, Brady MJ, Saltiel AR (1997) PTG, a protein phosphatase 1-binding protein with a role in glycogen metabolism. Science 275:1475–1478
2. Cheng C, Huang D, Roach PJ (1997) Yeast *PIG* genes: *PIG1* encodes a putative type I phosphatase subunit that interacts with the yeast glycogen synthase Gsy2p. Yeast 13:1–8
3. Bork P, Gibson T (1996) Applying motif and profile searches. Methods Enzymol 266:162–184
4. Schuler GD, Altschul SF, Lipman DJ (1991) A workbench for multiple alignment construction and analysis. Proteins Struct Funct Genet 9:180–190
5. Johnson DF, Moorhead G, Caudwell FB, Cohen P, Chen YH, Chen MX, Cohen PTW (1996) Identification of protein-phosphatase-1-binding domains on the glycogen and myofibrillar targetting subunits. Eur J Biochem 239:317–325
6. Svensson B, Larsen K, Gunnarson A (1986) Characterization of a glucoamylase G2 from *Aspergillus niger*. Eur J Biochem 154:497–502

7. Tanaka Y, Ashikari T, Nakamura N, Kiuchi N, Shibano Y, Amachi T, Yoshizumi H (1986) Comparison of amino acid sequences of three glucoamylases and their structure-function relationships. Agricultural Biol Chem 50:965–969
8. Altschul SF, Boguski MS, Gish W, Wootton JC (1994) Issues in searching molecular databases. Nat Genet 6:119–129
9. Svensson B, Jespersen H, Sierks MR, MacGregor EA (1989) Sequence homology between putative raw-starch binding domains from different starch-degrading enzymes. Biochem J 264:309–311
10. Bork P, Holm L, Sander C (1994) The immunoglobulin fold: Structural classification, sequence patterns and common core. J Mol Biol 242:309–320
11. Hubbard MJ, Cohen P (1993) On target with a new mechanism for the regulation of protein phosphorylation. Trends Biochem Sci 18:172–177
12. Wu J, Kleiner U, Brautigan DL (1996) Protein phosphatase type-1 and glycogen bind to a domain in the skeletal muscle regulatory subunit containing conserved hydrophobic sequence motifs. Biochemistry 35:13858–13864
13. Doherty MJ, Young PR, Cohen PTW (1996) Amino acid sequence of a novel protein phosphatase 1 binding protein (R5) which is related to the liver- and muscle-specific glycogen binding subunits of protein phosphatase 1. FEBS 399:339–343
14. Rost B (1996) Predicting one-dimensional protein structure by profile-based neural networks. Methods Enzymol 266:525–539
15. Egloff MP, Johnson DF, Moorhead G, Cohen PTW, Cohen P, Barford D (1997) Structural basis for the recognition of regulatory subunits by the catalytic subunit of protein phsophatase 1. EMBO J 16:1876–1887