Bioinformatics analysis of the structure - function relationship of glutamate dehydrogenase

Préambule

On peut considérer la première réaction d’assimilation de l’azote (sous forme d’ammoniac) par la glutamate déshydrogénase (GDH) comme un point d’entrée dans le métabolisme azoté. L’atome d’azote est à l’origine de la fonction α-aminée des acides aminés selon la réaction :

\[ \text{NH}_3^+ + \alpha\text{-cétoglutarate} + \text{NAD(P)H} + \text{H}^+ \rightleftharpoons \text{glutamate} + \text{NAD(P)}^+ \]

Il existe trois isoformes de GDH :

- la GDH EC 1.4.1.2 qui catalyse la réaction dans le sens de la désamination essentiellement
- la GDH EC 1.4.1.3 qui catalyse la réaction dans les deux sens
- la GDH EC 1.4.1.4 (GDH4) qui catalyse la réaction dans le sens de formation du glutamate

La GDH4 joue peut-être un rôle clé dans l’assimilation de l’azote. Or ce rôle n’a pas encore été démontré, notamment chez les plantes. Par ailleurs, on ne dispose d’aucune information concernant la structure de la GDH4.

La bioinformatique permet l’étude prospective de la relation structure - fonction de la GDH.
In term of taxonomy, the closest organism to higher plants (Eukaryota, Viridiplantae, Streptophyta) from which sequences of glutamate dehydrogenase EC 1.4.1.4 are known is an algae, Chlorella sorokiniana (Eukaryota, Viridiplantae, Chlorophyta).

Thus, the two sequences of this isoform from Chlorella sorokiniana were taken as the reference (Ref) subset in this study.

a. Overview of the search for amino acid sequences of the 3 isoforms of GDH

1. Go to the NCBI
2. Choose "ENTREZ"
3. Tape : "glutamate dehydrogenase OR GDH": 12293 hits (10/12/2010)
4. Click on : "Protein: sequence database"
5. Choose the option "Preview/Index". This allows the use of key-words with the option : "Add Term(s) to Query or View Index" and a boolean logical search ("AND" / "OR" / "NOT")

The number of hits is indicated on the right of the screen. To see the results ("Summary"), click on this number in your browser.

b. Classification of the 116 selected full GDHs amino acid sequences

116 non-redundant full GDH sequences were obtained from 83 organisms representing the three domains, Archaea, Bacteria and Eukaryota. These sequences were classified in 15 different subsets using the following criteria :

- the EC number
- the length of the polypeptide chain
- their Viridiplantae (higher plant) belonging or not

The following table gives the number of amino acid sequences of GDH for each subset (letter).

<table>
<thead>
<tr>
<th>Subset</th>
<th>Number of Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>411-470</td>
</tr>
<tr>
<td>2</td>
<td>503-558</td>
</tr>
<tr>
<td>3</td>
<td>1029-1106</td>
</tr>
<tr>
<td>4</td>
<td>1607-1651</td>
</tr>
</tbody>
</table>

| EC number | Viridiplantae | | | | | | not Viridiplantae | | | | total |
|-----------|--------------|---|---|---|---|---|---|---|---|---|---|---|
| 1.4.1.2   | L1 | 9 (A) | L2 | 7 (B) | L3 | 1 (C) | L4 | 17 |
| 1.4.1.3   | 1 (D) | L2 | 6 (E) | L3 | 7 (F) | L4 | 14 |
| 1.4.1.4   | 2 (Ref) | L2 | 15 (G) | L4 | 17 |
| not classified | L1 | 6 (H) | L2 | 13 (I1) | L3 | 18 (I2) | L4 | 14 (I3) | L1 | 5 (J) | L2 | 5 (K) | L3 | 7 (L) | L4 | 68 |
| total     | 16 | 2 | 73 | 12 | 5 | 8 | 116 |

The sequences of each subset were further aligned to obtain the 15 full consensus amino acids sequences.

c. Comparison of the 15 full consensus GDHs sequences

The schematic alignment of the full consensus sequences shows that GDH subunit is constituted of two or three regions:

- the N-terminal extension
- a common pattern to all consensus sequences corresponding to the central domain that contains the substrate and the nucleotide binding sites
- for large GDH (subsets C, K and L), the C-terminal extension

a. Unzip (untar) the text file containing the 15 full consensus sequences in FASTA format:

`FullConsSeq.zip` `FullConsSeq.tar`

b. Open the file with a text editor. Copy only the data beginning with a "">

c. Go to ClustalW (EBI). Paste the data into the window.
d. Select the appropriate matrix and parameters (below) and run the software.

- Matrix: Gonnet
- Gapopen: 1
- Gapext: 1
- Other parameters: default value

e. The results are returned: "*.aln" is the alignment.

d. Analysis of the central domain of GDH: the dinucleotide-binding motif

A $\beta\alpha\beta$ fold is found in the NAD(P)H-binding subdomain ($\beta_7 - \alpha_8 - \beta_8$).

This Rossmann fold begins with the motif $G^{313}AGNVA^{318}$ in the case of Ref.
However, the alignment indicates that the actual motives could be more complex. Such a higher complexity of the signature for the NAD(P)H-binding motif allows to discriminate more precisely the three isoforms.

This figure was generated using the software ESPript.

- Secondary structures indicated above the alignment were generated using as the template the bovine GDH3 complexed with NADPH and Glu (PDB # 1HWZ).
- Amino acid position indicated above the alignments is that of Ref (blue sequence).
- Plain red vertical boxes: amino acids identical for all consensus subsequences.
- Open red vertical boxes: amino acids whose homology between all consensus subsequences was greater than 60%.
- The letter "X" accounts for an amino acid whose identity level was less than 60% after the first alignment of full consensus sequences.
- The NAD(P)H-binding motif $G^{313}AGNVA^{318}$ (Ref) is indicated at the bottom of the frame with red circles.
Aldehyde dehydrogenase from *Vibrio harveyi* is one of the most NADP-specific.

The alignment of GDH from Ref and aldehyde DH shows that:

- there are three putative key residues for the binding of NADP(H) in Ref: Lys$^{202}$, Ser$^{205}$ (triangles) and Arg$^{248}$ (asterisk)

- the NAD(P)H-binding motif G$^{229}$SVGGG$^{234}$ of aldehyde DH is aligned with the motif G$^{266}$VLTGKG$^{272}$ of Ref (open circles)

Therefore, the latter is likely a second nucleotide-binding motif specific of GDH4.

This figure was generated using the software *ESprict*. 
f. Modelisation of the dinucleotide-binding motives and key residues of GDH4 with NADPH (NDP562) and Glu

A theoretical 3D structure of GDH4 from Ref was generated with the homology-modeling program ESyPred3D using as the template the structure of bovine GDH3 (PDB # 1HWZ).

The modelisation and the drawing of a putative structure of GDH4 was performed with the protein structure homology-modeling program DeepView (SwissPdb-Viewer v. 3.7).

Some interactions (plain lines) between the motif G313AGNVA318 or key residues and the coenzyme are indicated:

- NDP562AO3 - Gly313CA
- NDP562AO1 - Asn316ND2
- NDP562AO1 - Val317N
- NDP562AO2 - Gly244N
- NDP562NC4 - Thr285OG1
The distances between the protonated carbon atom of the nicotinamide moiety (NDP\textsuperscript{562}NC4) are too long for direct interactions with the motif G\textsuperscript{313}AGNVA\textsuperscript{318}.

However, this motif is stabilized by an internal H-bond Gly\textsuperscript{315}O - Ala\textsuperscript{318}N (dotted line).

Two distances (Glu\textsuperscript{557}OE2 - Lys\textsuperscript{166}NZ and Glu\textsuperscript{557}O - Lys\textsuperscript{190}NZ) are compatible with H-bond interactions between the enzyme and Glu.

The position of the motif G\textsuperscript{266}VLTGK\textsuperscript{272} is shown with the potential H-bond Lys\textsuperscript{166}NZ - Thr\textsuperscript{269}OG1.

\begin{tabular}{|l|l|}
\hline
\hline
\end{tabular}