Homology of the NifS family of proteins to a new class of pyridoxal phosphate-dependent enzymes

Christos Ouzounis and Chris Sander

Protein Design Group, European Molecular Biology Laboratory, Meyerhofstrasse 1, D-6900 Heidelberg, Germany

Received 19 March 1993

Iterative profile sequence analysis reveals a remote homology of peroxisomal serine-pyruvate aminotransferases from mammals to the small subunit of soluble hydrogenases from cyanobacteria, an isopenicillin N epimerase, the NifS gene products from bacteria and yeast, and the phosphoserine aminotransferase family. All members of this new class whose function is known are pyridoxal phosphate-dependent enzymes, yet they have distinct catalytic activities. Upon alignment, a lysine around position 200 remains invariant and is predicted to be the pyridoxal phosphate-binding residue. Based on the detected homology, it is predicted that NifS has also a pyridoxal phosphate-dependent serine (or related) aminotransferase function associated with nitrogen economy and/or protection during nitrogen fixation.

Serine-pyruvate aminotransferase; Soluble hydrogenase small subunit; NifS; Protein sequence analysis; Function prediction by sequence homology

1. INTRODUCTION

Aminotransferases seem to be an extensive group of functionally related enzymes derived from different phylogenetic origins. Since catalysis is usually performed with the help of pyridoxal-phosphate covalently bound to a lysine residue [1], and not with a particular constellation of amino acids, aminotransferase classes are only related by convergent evolution to a common functional property. It is therefore interesting to investigate by sequence analysis the evolutionary relationships among aminotransferases and identify members that are divergently related.

Such an attempt has previously been made and sequence analysis reveals three classes of aminotransferases, plus a single phosphoserine aminotransferase family (A. Bairoch, unpublished observations). Class I is represented by the prototype aspartate aminotransferase [1] and the related aromatic aminotransferases. Class II includes glycine aminotransferases and related enzymes. Class III includes ornithine aminotransferases and related enzymes. Finally, phosphoserine aminotransferases define a separate family [2]. Here we examine the sequences of two other serine-pyruvate aminotransferases [3–5] that have not been classified under the above scheme, and we show that they belong to a fourth class of fifteen known enzymes, some of which are associated with serine transamination. This class includes the above mentioned family of phosphoserine aminotransferases, and other enzymes of related or unknown function.

2. MATERIALS AND METHODS

2.1. Profile sequence analysis

Evidence for the remote sequence similarities identified here comes from profile searches [6] and multiple alignments [7]. The following iterative procedure was used: a profile is first constructed, then a database search is performed, and any additional significant hits are subsequently aligned and a new profile is generated. The criteria for significance were (i) that the normalized score should be greater than 6.00 (six standard deviations above the mean) and (ii) that the identified protein should be longer than 300 residues. The procedure terminates when no additional significant hits, using the above criteria, can be identified [8]. The numerical details of the present database searches are shown in Table I.

2.2. Multiple sequence alignment

Multiple alignments were generated using the program PILEUP of the GCG sequence analysis software package version 7.1. [9]. Gap penalties were set to 3.0 for opening and 0.1 for elongation. This multiple alignment method is based on the pairwise dynamic programming alignment [10] and subsequent clustering of a distance matrix using the unweighted pair-group method using arithmetic averages (UPGMA) clustering algorithm [7,11]. The amino acid composition matrix was the normalized Dayhoff matrix [12]. Information about the physical properties of the proteins discussed here is given in Table II.

3. RESULTS AND DISCUSSION

3.1. A new class of serine aminotransferases and related enzymes

Profile searches starting with the two known peroxisomal serine-pyruvate aminotransferases from mammals, reveal a homology to four other families of proteins (see Table I). These families are (a) the small subunit of soluble hydrogenases from cyanobacteria,
(b) isopenicillin N epimerase, (c) the NifS family from bacteria and yeast and (d) phosphoserine aminotransferases (for references, see Table II).

Interestingly, in an independent study it was shown that peroxisomal serine-pyruvate aminotransferases from mammals are homologous to NifS [27]. However, the other homologies presented here were not reported.

A dendrogram that represents graphically the sequence similarity relationships within the class has been constructed and it is evident that the fifteen proteins fall into five families on the basis of sequence similarity (Fig. 1). Sequence identity is below 25% over 400 residues for sufficiently distant members within the class. The detection of such remote relationships becomes feasible only through the use of multiple sequence alignment and profile sequence analysis [28].

A multiple alignment of representative members of the new class is given in Fig. 2. The length of the aligned sequences is 400 residues, comparable with the average length of members of the new class. The relative constancy of sequence length upon alignment is an additional indication for an evolutionary relationship in such cases. The alignment of these remotely related proteins reduces the set of conserved residues to very few positions, suitable candidates for mutagenesis. Few residues are invariant (K208, G215, and T267) and these could possibly participate in the formation of the active site. Moreover, residues D181, D339 and R387 although not conserved in all sequences have neighboring conserved residues that are possibly functionally equivalent (Fig. 2).

Based on secondary structure predictions and comparison with other aminotransferases, it has already been suggested that lysine 208 is the covalent ligand for pyridoxal phosphate, in the case of serine-pyruvate aminotransferases [3]. For none of the other members of the class an active site residue has been predicted or identified so far. Our secondary structure predictions [29] indicate that the class might have an overall alpha-beta topology, possibly similar to the class I aminotransferases [1]. Moreover, lysine 208 is predicted with high accuracy to be located in a loop region between two strands (Fig. 3).

The structural and functional properties of each of these families and the significance of the remote sequence similarities within this class are discussed below:

(a) The sequence similarity between mammalian peroxisomal serine-pyruvate aminotransferases and the small subunit of soluble hydrogenases is around 30%. This number is five percentage points above the level for structural homology and clearly indicates evolutionary divergence [30]. Despite this relatively high sequence

### Table I

<table>
<thead>
<tr>
<th>SWISSPROT</th>
<th>1st search</th>
<th>2nd search</th>
<th>3rd search</th>
<th>4th search</th>
<th>Len</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Z</td>
<td>T</td>
<td>Z</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td><strong>1st set</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>392</td>
</tr>
<tr>
<td>SpyA_Human</td>
<td>85.20</td>
<td>472.87</td>
<td>54.77</td>
<td>181.72</td>
<td>43.72</td>
</tr>
<tr>
<td>SpyA_Rat</td>
<td>82.86</td>
<td>473.07</td>
<td>55.09</td>
<td>181.99</td>
<td>41.66</td>
</tr>
<tr>
<td><strong>+ 2nd set</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>389</td>
</tr>
<tr>
<td>Dhss_Sympt</td>
<td>19.82</td>
<td>144.97</td>
<td>50.88</td>
<td>169.02</td>
<td>39.76</td>
</tr>
<tr>
<td>Dhss_Ancay</td>
<td>18.71</td>
<td>139.29</td>
<td>50.28</td>
<td>167.12</td>
<td>38.95</td>
</tr>
<tr>
<td>CefD_Strcl</td>
<td>7.62</td>
<td>85.99</td>
<td>34.09</td>
<td>124.67</td>
<td>26.98</td>
</tr>
<tr>
<td><strong>+ 3rd set</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>418</td>
</tr>
<tr>
<td>NifS_KlepN</td>
<td>6.90</td>
<td>48.10</td>
<td>29.65</td>
<td>80.71</td>
<td>34.69</td>
</tr>
<tr>
<td>NifS_Anap</td>
<td>6.49</td>
<td>47.12</td>
<td>31.61</td>
<td>84.93</td>
<td>35.66</td>
</tr>
<tr>
<td>FprP_Rabat</td>
<td>6.69</td>
<td>45.77</td>
<td>73.97</td>
<td>67.00</td>
<td>28.65</td>
</tr>
<tr>
<td><strong>+ 4th set</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>400</td>
</tr>
<tr>
<td>NifS_Azoch</td>
<td>21.56</td>
<td>64.53</td>
<td>32.03</td>
<td>58.00</td>
<td>38.80</td>
</tr>
<tr>
<td>NifS_Azovi</td>
<td>26.06</td>
<td>74.02</td>
<td>37.54</td>
<td>66.11</td>
<td>35.66</td>
</tr>
<tr>
<td>NifG_Yeast</td>
<td>13.68</td>
<td>54.06</td>
<td>25.53</td>
<td>51.96</td>
<td>35.66</td>
</tr>
<tr>
<td>Serc_Ecoli</td>
<td>11.90</td>
<td>43.31</td>
<td>36.41</td>
<td>61.40</td>
<td>35.44</td>
</tr>
<tr>
<td>Serc_Salgi</td>
<td>11.76</td>
<td>43.05</td>
<td>35.44</td>
<td>60.11</td>
<td>35.44</td>
</tr>
<tr>
<td>Serc_Yeren</td>
<td>11.33</td>
<td>42.94</td>
<td>32.22</td>
<td>55.75</td>
<td>35.44</td>
</tr>
</tbody>
</table>
Table II

Information concerning the proteins discussed in the text

SWISSPROT identifiers are used, except Nifs/Lacde which is taken from PIR database; this convention is also followed in the figures. ACCESS is the accession number. LEN is the length of the protein, MW the calculated molecular weight, PLP is the predicted pyridoxal phosphate-binding lysine, with the exception of the serine-pyruvate aminotransferases [3]. REF is the original sequencing report.

<table>
<thead>
<tr>
<th>SWISSPROT</th>
<th>Access</th>
<th>Protein name</th>
<th>Species name</th>
<th>LEN</th>
<th>MW</th>
<th>PLP</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SpyA_Human</td>
<td>P21549</td>
<td>Serine-Pyruvate Aminotransferase</td>
<td>Homo sapiens</td>
<td>392</td>
<td>43,010</td>
<td>209</td>
<td>[5]</td>
</tr>
<tr>
<td>SpyA_Rat</td>
<td>P09139</td>
<td>Serine-Pyruvate Aminotransferase</td>
<td>Rattus norvegicus</td>
<td>414</td>
<td>45,834</td>
<td>231</td>
<td>[3,4]</td>
</tr>
<tr>
<td>Dhss_Anacy</td>
<td>P16421</td>
<td>Soluble Hydrogenase 42 kDa Subunit</td>
<td>Anabaena cylindrica</td>
<td>383</td>
<td>41,155</td>
<td>194</td>
<td>[14]</td>
</tr>
<tr>
<td>Dhss_Synpl</td>
<td>P14776</td>
<td>Soluble Hydrogenase Small Subunit</td>
<td>Synechococcus sp. PCC 6716</td>
<td>384</td>
<td>40,876</td>
<td>194</td>
<td>[15]</td>
</tr>
<tr>
<td>CefD_Strcl</td>
<td>P18549</td>
<td>Isopenicillin N Epimerase</td>
<td>Streptomyces clavuligerus</td>
<td>397</td>
<td>43,366</td>
<td>216</td>
<td>[16]</td>
</tr>
<tr>
<td>Nifs_Klep</td>
<td>P05344</td>
<td>NIFS Protein</td>
<td>Klebsiella pneumoniae</td>
<td>397</td>
<td>43,254</td>
<td>199</td>
<td>[17]</td>
</tr>
<tr>
<td>Nifs_Anasp</td>
<td>P12623</td>
<td>NIFS Protein</td>
<td>Anabaena sp. PCC 7120</td>
<td>400</td>
<td>43,657</td>
<td>201</td>
<td>[18]</td>
</tr>
<tr>
<td>Nifs_Azoch</td>
<td>P10658</td>
<td>Probable Phosphoserine Aminotransferase</td>
<td>Oryctolagus cuniculus</td>
<td>370</td>
<td>50,621</td>
<td>200</td>
<td>[19]</td>
</tr>
<tr>
<td>Nifs_Azovi</td>
<td>P23120</td>
<td>NIFS Protein</td>
<td>Azotobacter chroococcum</td>
<td>396</td>
<td>43,296</td>
<td>199</td>
<td>[20]</td>
</tr>
<tr>
<td>Nifs_Lacde</td>
<td>P05341</td>
<td>NIFS Protein</td>
<td>Azotobacter vinelandii</td>
<td>402</td>
<td>43,598</td>
<td>202</td>
<td>[21]</td>
</tr>
<tr>
<td>Nifs_Yeast</td>
<td>P25374</td>
<td>NIFS-Like 54.5 kDa Protein</td>
<td>Saccharomyces cerevisiae</td>
<td>497</td>
<td>54,493</td>
<td>299</td>
<td>[22]</td>
</tr>
<tr>
<td>Nifs_Lacte</td>
<td>S16047</td>
<td>NIFS-Like Protein</td>
<td>Lactobacillus delbrueckii</td>
<td>355</td>
<td>38,792</td>
<td>191</td>
<td>[23]</td>
</tr>
<tr>
<td>Sere_Ecoli</td>
<td>P23721</td>
<td>Phosphoserine Aminotransferase</td>
<td>Escherichia coli</td>
<td>362</td>
<td>39,686</td>
<td>198</td>
<td>[24]</td>
</tr>
<tr>
<td>Sere_Salgl</td>
<td>P17902</td>
<td>Phosphoserine Aminotransferase</td>
<td>Salmonella gallinarum</td>
<td>362</td>
<td>39,855</td>
<td>198</td>
<td>[25]</td>
</tr>
<tr>
<td>Sere_Yeren</td>
<td>P19689</td>
<td>Phosphoserine Aminotransferase</td>
<td>Yersinia enterocolitica</td>
<td>361</td>
<td>40,162</td>
<td>197</td>
<td>[26]</td>
</tr>
</tbody>
</table>

Similarity, this relationship has not been reported before. There are two types of hydrogenases in nitrogen-fixing organisms: (i) the soluble hydrogenase [31], present both in undifferentiated cells and heterocysts (discussed here) and (ii) the uptake hydrogenase [32], which is membrane-bound and present only in heterocysts [33]. It is not known whether the small subunit of soluble hydrogenase is a pyridoxal phosphate-dependent enzyme. From the alignment reported here, it is predicted that lysine 194 (K208 in Fig. 2) forms the cofactor-binding active site.

(b) Isopenicillin N epimerase [16] is identified as homologous to serine-pyruvate aminotransferases by the first profile database search. The level of sequence identity with either mammal aminotransferase is within the 20–22% range. Isopenicillin N epimerase has been shown to contain one molecule of pyridoxal phosphate per monomer [34]. Based on the detected homology, the cofactor-binding residue is predicted to be lysine 216 (K208 in Fig. 2).

(c) The enzymes responsible for nitrogen fixation are encoded by a cluster of at least 17 nitrogen fixation genes (nif genes) in the Klebsiella chromosome [35]. Other bacteria have a similar gene organization [17, 36]. These microorganisms, including cyanobacteria, have the ability to fix atmospheric nitrogen by reduction of dinitrogen to ammonia [37].

Dinitrogenase, the Mo-Fe component of the nitrogenase system, is encoded by nifD and nifK, while dinitrogen reductase, the Fe-protein, is encoded by nifH [38]. These enzymes produce ammonia which is incorporated into glutamate, and subsequently to glutamine and other organic nitrogen compounds by transaminase or transamidase reactions [38].

Other nif genes (nifL and nifA) code for proteins that regulate expression in this particular gene cluster [35]. Finally, some nif genes code for proteins involved in synthesis and activation of specific nitrogenase components and cofactors as well as oxidation-reduction reactions and electron transfer [38]. Yet, for some genes, including nifS, the function remains unknown. It has been proposed that NifS is essential for the formation of the nitrogenase enzyme complex [39], but its exact catalytic activity has not yet been determined. NifS appears to play a role in Fe-protein production in Azoto-

Fig. 1. A dendrogram representing sequence similarity relationships during multiple alignment construction. It is evident that the fifteen sequences fall into five families. Information about sequences can be found in Table II.

161
bacter vinelandii, but not in Klebsiella pneumoniae [40], although it is required for MoFe-protein formation.

Recently, with the sequencing of the complete chromosome III of yeast and the identification of a nif/S homologue in a eukaryotic, non-nitrogen fixing organism [22], interest in the function of this gene has increased. The present observation might contribute to mutant construction and functional analysis for the Nifs family [41].

(d) Phosphoserine aminotransferases catalyze the third and last step of serine biosynthesis from 3-phosphoglycerate [24]. It is known that this transamination reaction also requires pyridoxal phosphate as a cofactor. The enzyme and the corresponding pathway seem to be widespread from bacteria to man: a mammal progesterone-induced gene [19] was shown to code for a protein homologous to phosphoserine aminotransferases from bacteria [42].

3.2. A possible function for Nifs: a serine aminotransferase participating in nitrogen fixation?

It is puzzling that Nifs is homologous with protein families having such different functions. However, for those proteins whose function is known, it is evident that they all are pyridoxal phosphate-dependent enzymes. In fact, six of these proteins are (phospho)serine aminotransferases; with the exception of isopenicillin N epimerase, the exact function of the remaining members is not known. It is therefore plausible that Nifs is an aminotransferase, catalyzing the interconversion of organic nitrogen compounds and/or the activation and regulation of nitrogenase activity.

There can be at least two, not mutually exclusive, reasons why amino acid interconverting processes are needed during nitrogen fixation: first, it is crucial to provide fast and reliable synthesis of specific amino acids required for the translation of other nif genes including nitrogenase, and second, more important, it is essential to redistribute and store nitrogen in other compounds without back-inhibiting ammonia production.

Interestingly, the identification of a serine acetyltransferase (nifP gene), a key enzyme in cysteine biosynthesis, in the nif cluster of Azotobacter chroococcum [20], points out the requirement of enzymes associated with amino acid biosynthesis, for the optimal expression of nitrogenase activity.

The role of aminotransferase reactions for the economy of nitrogen is well-documented [43]. Nitrogen fixation over-produces amino acids like glutamate and aspartate [44]. Need for interconversion to other, less abundant, amino acids is then essential. Aminotransferases provide the route for redistribution of amino acid nitrogen, using transamination as a mechanism of amino acid synthesis or degradation [45]. The fact that the serine to glycine interconversion occurs in mitochondria and peroxisomes (organelles of a prokaryotic origin) by the same gene product in mammals [4] provides additional support for the prokaryotic origin of serine-pyruvate aminotransferases.

The role of hydrogenases in nitrogen fixation is not completely understood: nitrogen fixation is accompanied by a variable amount of proton reduction and molecular hydrogen formation [37, 46]. Hydrogenases catalyze the consumption of H2, a reaction that is important both for the removal of hydrogen during ammonia production and the reduction of energy waste due to hydrogen production. It is known that nitrogenase components are extremely sensitive to oxygen [47]. It is remarkable that soluble hydrogenase is active during nitrogen fixation of nitrogen [38], due to hydrogen production. It is known that nitrogenase activity [37]. However, its exact molecular function remains unknown.

Based on the present study, we propose that this class of known enzymes has a common evolutionary
origin and a divergent history of 2.5 billion years dating back to the origins of cyanobacteria [48].

Acknowledgements: Expert advice and help by Peer Bork (EMBL) has been invaluable. Information by Amos Bairoch (Univ. of Geneva) on aminotransferases is gratefully acknowledged.

REFERENCES