Volume 322, number 2, 159–164 FEBS 12421 May 1993

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analysis reveals three classes of aminotransferases, plus a single phosphoserine aminotransferase family (A. Bairoch, unpublished observations).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 other 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.

Correspondence address: C. Ouzounis, Protein Design Group, European Molecular Biology Laboratory, Meyerhofstrasse 1, D-6900 Heidelberg, Germany. Fax: (49) (6221) 387 306.

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. [10]. 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 [11] and subsequent clustering of a distance matrix using the unweighted pair-group method using arithmetic averages (UPGMA) clustering algorithm [7,12]. The amino acid comparison matrix was the normalized Dayhoff matrix [13]. 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

Iterative profile sequence analysis for the identification of remote homologues

A profile is generated for the entries of set *i* and scans the database in the ith search. The set *i* + *1* contains the significant hits that are included in the generation of another profile along with the members of set *i*. The procedure terminates when no additional significant hits are found. Z is the normalized standard score, and T is the total score of the alignment. Note that the inclusion of a sequence in the profile does not always result to an increased score, as the searches strongly depend on the presence of homologous sequences. Len is profile length. Sequence identifiers are taken from the SWISSPROT database, release 23 [9].

SWISSPROT	1st search		2nd search		3rd search		4th search		Len
	Z	T	Z	T	Z	T	Z	T	
1st set									392
Spya_Human	85.20	472.87	54.77	181.72	43.72	108.04	27.50	51.47	
Spya_Rat	82.86	473.07	53.09	181.99	41.66	106.73	26.64	51.55	
+ *2nd set*									389
Dhss_Synpl	19.82	144.97	50.88	169.02	39.76	99.19	24.70	47.14	
Dhss_Anacy	18.71	139.29	50.28	167.12	38.95	97.48	23.53	45.50	
Cefd_Strcl	7.62	85.99	34.09	124.67	26.98	75.38	17.52	37.92	
+ *3rd set*									418
Nifs_Klepn			6.90	48.10	29.65	80.71	34.69	61.76	
Nifs_Anasp			6.49	47.12	31.61	84.93	35.66	63.33	
Epip_Rabit			6.69	45.77	23.97	67.00	28.65	51.62	
+ *4th set*									400
Nifs_Azoch					21.56	64.53	32.03	58.00	
Nifs_Azovi					26.06	74.02	37.54	66.11	
Nfsl_Yeast					13.68	54.06	25.53	53.96	
Serc_Ecoli					11.90	43.31	36.41	61.40	
Serc_Salgl					11.76	43.05	35.44	60.11	
Serc_Yeren					11.73	42.94	32.22	55.75	

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 aminotransferaseses [3]. REF is the original sequencing report.

SWISSPROT	Access	Protein name	Species name	LEN	MW	PLP	Ref.
Spya_Human	P21549	Serine-Pyruvate Aminotransferase	*Homo sapiens*	392	43,010	209	[5]
Spya_Rat	P09139	Serine-Pyruvate Aminotransferase	*Rattus norvegicus*	414	45,834	231	[3,4]
Dhss_Anacy	P16421	Soluble Hydrogenase 42 kDa Subunit	*Anabaena cylindrica*	383	41,155	194	[14]
Dhss_Synpl	P14776	Soluble Hydrogenase Small Subunit	*Synechococcus* sp. PCC 6716	384	40,876	194	[15]
Cefd_Strcl	P18549	Isopenicillin N Epimerase	*Streptomyces clavuligerus*	397	43,366	216	[16]
Nifs_Klepn	P05344	NIFS Protein	*Klebsiella pneumoniae*	397	43,254	199	[17]
Nifs_Anasp	P12623	NIFS Protein	*Anabaena* sp. PCC 7120	400	43,657	201	[18]
Epip_Rabit	P10658	Probable Phosphoserine Aminotransferase	*Oryctolagus cuniculus*	370	50,621	200	[19]
Nifs_Azoch	P23120	NIFS Protein	*Azotobacter chroococcum*	396	43,296	199	[20]
Nifs_Azovi	P05341	NIFS Protein	*Azotobacter vinelandii*	402	43,598	202	[21]
Nfsl_Yeast	P25374	NIFS-Like 54.5 kDa Protein	*Saccharomyces cerevisiae*	497	54,493	299	[22]
Nifs/Lacde	S16047	NIFS-Like Protein	*Lactobacillus delbrueckii*	355	38,792	191	[23]
Serc_Ecoli	P23721	Phosphoserine Aminotransferase	*Escherichia coli*	362	39,868	198	[24]
Serc_Salgl	P17902	Phosphoserine Aminotransferase	*Salmonella gallinarum*	362	39,855	198	[25]
Serc_Yeren	P19689	Phosphoserine Aminotransferase	*Yersinia enterocolitica*	361	40,162	197	[26]

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 coenzyme-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 coenzyme-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 dinitrogenase 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-*

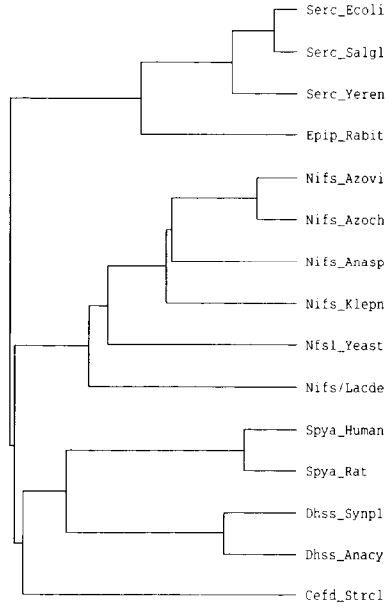

Fig. 1. A dendrogram representing sequence similarity relationships during multiple alignment construction. It is evident that the fifteen sequences fall into five families. Information about sequences can be found in Table II.

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 *nifS* 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 H_2, 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, yet independent of, nitrogen-fixing activity [37]. However, its exact molecular function remains unknown.

Based on the present study, we propose that this class of fifteen known enzymes has a common evolutionary

```
                  201    !         220
Serc_Ecoli        VIYAGAQKNI.GPAGLTIVI
Serc_Salgl        VIYAGAQKNI.GPAGLTLVI
Serc_Yeren        VIYAGAQKNI.GPAGLTVVI
Epip_Rabit        VIFAGAQKNV.GAAGVTVVI
Nifs_Azoch        MLSLSGHK.LH.RKGVGVLY
Nifs_Azovi        MLSLCGHK.LHAPKGVGVLY
Nifs_Anasp        MLTISGHK.IHAPKGIGALY
Nifs_Klepn        MLSCSAHK.FHGPKGVGCLY
Nfs1_Yeast        LLSISSHK.IYGPKGIGAIY
Nifs/Lacde        MMSFSSHK.FHGPRGIGILY
Spya_Human        ILYSGSQKALNAPPGTSLIS
  Spya_Rat        ILYSGSQKVLNAPPGISLIS
Dhss_Anacy        VVASGSQKGYMIPPGLGFVS
Dhss_Synpl        VVGSGSQKGYMIPPGLAFVS
Cefd_Strcl        FYAGSGHKWLLAPTGVGFLH

 Consensus        VL  G QK L  P GVG V
 Prediction       EEEE    LLLLLLL EEEE
 Reliability      55413433565776515676
```

Fig. 3. Multiple alignment and secondary structure prediction around the predicted covalent ligand for pyridoxal phosphate (lysine 208). Consensus as calculated in Fig. 2. Predicted secondary structure for the class (E: strand, L: loop) at 82% accuracy; reliability index of prediction (scale 0–9) [29].

←

Fig. 2. A multiple alignment constructed for all fifteen members of the class. Sequence numbering is also given. Each line contains 100 amino acids. The consensus line is calculated from at least eight most frequently occurring residues at a particular position, taking into account amino acid similarity values. Boxed residues represent similarity with the consensus. Residues conserved in at least sequences are marked with an asterisk. Invariant residues are marked with an exclamation point. Neighboring residue positions that might be functionally conserved and are present in all fifteen sequences are marked with numbers. Alternative corresponding residues for R386 (3) are marked with 3′ and 3″. Figure generated by PRETTYPLOT (courtesy of Peter Rice, EMBL).

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

[1] Kirsch, J.F., Eichele, G., Ford, G.C., Vincent, M.G. and Jansonius, J.N. (1984) J. Mol. Biol. 174, 497–525.
[2] Bairoch, A. (1992) Nucleic Acids Res. 20, 2013–2018.
[3] Oda, T., Miyajima, H., Suzuki, Y. and Ichiyama, A. (1987) Eur. J. Biochem. 168, 537–542.
[4] Oda, T., Funai, T. and Ichiyama, A. (1990) J. Biol. Chem. 265, 7513–7519.
[5] Nishiyama, K., Berstein, G., Oda, T. and Ichiyama, A. (1990) Eur. J. Biochem. 194, 9–18.
[6] Gribskov, M., McLachlan, M. and Eisenberg, D. (1987) Proc. Natl. Acad. Sci. USA 84, 4355–4358.
[7] Higgins, D.G., Bleasby, A.J. and Fuchs, R. (1992) Comput. Appl. Biosci. 8, 189–191.
[8] Gribskov, M. (1992) Gene 119, 107–111.
[9] Bairoch, A. and Boeckmann, B. (1992) Nucleic Acids Res. 20, 2019–2022.
[10] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
[11] Needleman, S.B. and Wunsch, C.D. (1970) J. Mol. Biol. 48, 443–453.
[12] Feng, D.-F. and Doolittle, R.F. (1987) J. Mol. Evol. 25, 351–360.
[13] Gribskov, M. and Burgess, R.R. (1986) Nucleic Acids Res. 14, 6745–6763.
[14] Ewart, G.D., Reed, K.C. and Smith, G.D. (1990) Eur. J. Biochem. 187, 215–223.
[15] Van der Oost, J., van Walraven, H.S., Bogerd, J., Smit, A.B., Ewart, G.D. and Smith, G.D. (1989) Nucleic Acids Res. 17, 10098–10098.
[16] Kovacevic, S., Tobin, M.B. and Miller, J.R. (1990) J. Bacteriol. 172, 3952–3958.
[17] Beynon, J., Ally, A., Cannon, M., Cannon, F., Jacobson, M., Cash, V. and Dean, D. (1987) J. Bacteriol. 169, 4024–4029.
[18] Mulligan, M.E. and Haselkorn, R. (1989) J. Biol. Chem. 264, 19200–19207.
[19] Misrahi, M., Atger, M. and Milgrom, E. (1987) Biochemistry 26, 3975–3982.
[20] Evans, D.J., Jones, R., Woodley, P.R., Wilborn, J.R. and Robson, R.L. (1991) J. Bacteriol. 173, 5457–5469.
[21] Jacobson, M.R., Brigle, K.E., Bennett, L.T., Setterquist, R.A., Wilson, M.S., Cash, V.L., Beynon, J., Newton, W.E. and Dean, D.R. (1989) J. Bacteriol. 171, 1017–1027.
[22] Oliver, S.G. et al., (1992) Nature 357, 38–46.
[23] Leong-Morgenthaler P.M., Oliver, S. and Söll, D. (1991) unpublished.
[24] Duncan, K. and Coggins, J.R. (1986) Biochem. J. 234, 49–57.
[25] Griffin, H.G. (1990) Nucleic Acids Res. 18, 4260–4260.
[26] O'Gaora, P., Maskell, D., Coleman, D., Cafferkey, M. and Dougan, G. (1989) Gene 84, 23–30.
[27] Mehta, P.K. and Christen, P. (1993) Eur. J. Biochem. 211, 373–376.
[28] Gribskov, M., Lüthy, R. and Eisenberg, D. (1990) Methods Enzymol. 183, 146–159.
[29] Rost, B. and Sander, C. (1993) Proc. Natl. Acad. Sci. USA, in press.
[30] Sander, C. and Schneider, R. (1991) Proteins 9, 56–68.
[31] Houchins, J.P. and Burris, R.H. (1981) Plant Physiol. 68, 717–721.
[32] Tel-Or, E.L., Luijk, L.W. and Packer, L. (1978) FEBS Lett. 78, 49–52.
[33] Haselkorn, R. (1978) Annu. Rev. Plant Physiol. 29, 319–344.
[34] Usui, S. and Yu, C.-A. (1989) Biochim. Biophys. Acta 999, 78–85.
[35] Gussin, G.N., Ronson, C.W. and Ausubel, F.M. (1986) Annu. Rev. Genet. 20, 567–591.
[36] Mulligan, M.E., Buikema, W.J. and Haselkorn, R. (1988) J. Bacteriol. 170, 4406–4410.
[37] Fay, P. (1992) Microbiol. Rev. 56, 340–373.
[38] Orme-Johnson, W.H. (1985) Annu. Rev. Biophys. Biophys. Chem. 14, 419–459.
[39] Golden, J.W., Carrasco, C.D., Mulligan, M.E., Schneider, G.J. and Haselkorn, R. (1988) J. Bacteriol. 170, 5034–5041.
[40] Harris, G.S., White, T.C., Flory, J.E. and Orme-Johnson, W.H. (1990) J. Biol. Chem. 265, 15909–15919.
[41] Kennedy, C. and Dean, D. (1992) Mol. Gen. Genet. 231, 494–498.
[42] Van der Zel, A., Lam, H.-M. and Winkler, M. (1989) Nucleic Acids Res. 17, 8379–8379.
[43] Metzler, D.E. (1977) Biochemistry (chapter 14), Academic Press, New York.
[44] Miflin, B.J. and Lea, P.J. (1976) Trends Biochem. Sci. 1, 103–106.
[45] Matthews, C.K. and van Holde, K.E. (1990) Biochemistry (chapter 20), Benjamin/Cummings Publishing Company, Redwood City.
[46] Robson, R.L. and Postgate, J.R. (1980) Annu. Rev. Microbiol. 34, 183–207.
[47] Mortenson, L.E. and Thorneley, R.N.F. (1979) Annu. Rev. Biochem. 48, 387–418.
[48] Schopf, J.W. and Walter, M.R. (1982) in: The biology of cyanobacteria (Carr, N.G. and Whitton, B.A., eds.) pp. 543–564, Blackwell, Oxford.