Vol. 171, No. 3, 1990 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
September 28, 1990 Pages 1319-1325

SEQUENCE SIMILARITIES BETWEEN TRYPTOPHAN SYNTHASE β SUBUNIT AND OTHER PYRIDOXAL-PHOSPHATE-DEPENDENT ENZYMES

Peer Bork and Klaus Rohde

Academy of Sciences of the GDR, Central Institute of Molecular Biology, Department of Biomathematics, Berlin 1115, Germany

Received August 20, 1990

Summary: On the basis of 8 tryptophan synthase β subunits (EC 4.2.1.20) consensus patterns were constructed comprising two conserved motifs. Screening of the SWISSPROT protein sequence database with these patterns indicates similarities with O-acetylserine sulfhydrolases (EC 4.2.99.8), threonine synthases (EC 4.2.99.2), L- and D-serine dehydratases (EC 4.2.1.13/ EC 4.2.1.14) and threonine dehydratases (EC 4.2.1.16). Using multiple alignment procedures the similar regions could be extended. In connection with their pyridoxal-phosphate-binding-capacity and their positions in biochemical pathways evolutionary relationships among these enzymes are discussed. © 1990 Academic Press, Inc.

With the progress in sequencing techniques more and more primary structures of pyridoxal-phosphate-dependent enzymes become available. This offers the chance to get information about tertiary structures, functional mechanisms and evolutionary relationships. Since pyridoxal - phosphate is a cofactor of different enzymes like transaminases, decarboxylases, phosphorylases, dehydratases, synthases, sulfhydrolases etc. and since pyridoxal-phosphate is able to catalyse many of the respective reactions alone (without enzymes) the structures and sequences of these proteins can be expected to vary to a great extent and there is no reason for a common ancestor. This is confirmed by high resolution X-ray crystallography of glycogen phosphorylase A and B (1,2) some aspartate aminotransferases (mitochondrial isoenzyme from chicken heart (3), cytosolic isoenzymes from pig (4) and chicken heart (5)) and tryptophan synthase (6) for which no structural homology could be obtained. Other than, for instance, in nucleotide binding proteins (7) the binding sites (around the essential lysine) are different in the published tertiary structures (1,5,6).

0006-291X/90 $1.50
Copyright © 1990 by Academic Press, Inc.
All rights of reproduction in any form reserved.

Pyridoxal-phosphate-dependent enzymes are involved in many biochemical pathways, but an accumulation is found in anabolic reactions of the aspartate family and in connected pathways. Some homologies among pyridoxal-phosphate-dependent enzymes of the biosynthetic pathways of the aspartate family could be identified (8). Despite different functions they seem to have similar structures. Recently, distant homologies were found also between some aminotransferases (9), but the enzymes of this family act in a completely different way.

As a further application of our simple property pattern search method (10), which was also sensitive in identifying different nucleotide binding sites (11), we propose here a distant homology between tryptophan synthase β subunit and other pyridoxal-phosphate-dependent enzymes. Homologies among this family of pyridoxal-phosphate-dependent enzymes allow structural and functional insights in the mechanisms of all these enzymes and provide some hints related to the evolution of biochemical pathways.

Materials and Methods

By means of FASTP (12) the sequences most similar to tryptophan synthase β subunit of salmonella typhimurium (with known structure) could be identified in SWISSPROT protein sequence database (release 12.0, containing 12305 sequences). Significant matches were given by tryptophan synthase β subunits from other species and the equivalent part of the uncleaved yeast enzyme. The 8 sequences found in this way were aligned using our program MULTIS (13) which allows to compare up to 15 protein sequences now.

In tryptophan synthase pyridoxal-phosphate is covalently bound to an essential lysine of the β subunit (or to the equivalent region of the uncleaved enzymes). A second conserved region is a glycine rich turn which ligates through H-bonds the phosphate group of pyridoxal-phosphate (6). For these regions property patterns were constructed using PAT (14). This procedure leads to rows of amino acids in each position of a sequence segment (fig.1b). With these patterns the SWISSPROT database was screened.

Since our simple property pattern procedure is able to recognize also distantly related proteins (10), for multiple alignment of protein sequences with a low degree of amino acid identity we have developed an additional program PULIN, which carries out the multiple alignment via an iterated profile alignment (15-17) using the Dayhoff PAM250 matrix (18). We used PULIN to align those proteins, for which some similar regions were detected by our property pattern method.

Results

When the SWISSPROT was screened with our property pattern comprising the two motifs described above in addition to the

a)

codes	proteins	lysine-motif			glycine rich turn		
		pos	m	sequence segments	pos	m	sequence segment
CYSK$ECOLI	O-acetylserine sulfhydrolase	25	3	NGRILAKvESRNP-SFSvKCRiGAN	167	3	GQVDVfIAgVGTGGTLTGVsRYI
CYSK$SALTY	:	25	3	NGRILAKvESRNP-SFSvKCRiGAN	167	3	GQVDVfIsgVGTGGTLTGVTRYI
SDHD$ECOLI	D-serine dehydratase	100	2	SGQLLLKKDSHLPISGSIKaRGGIy	269	1	DNPLFVYLPCGvGGGPGGVAFGL
SDHL$RAT	L-serine dehydratase	24	1	GTSVFLKMDSSQP-SGSFKIRGIGh	194	0	AKPGAIVLSVGGGGLLCGVVQGL
THD1$ECOLI	threonine dehydratase	45	0	DNVILVKREDRQP-VHSFKLRGAYA	178	1	AHLDRVFVPVGGGGLAAGVAVlI
THD2$ECOLI	:	41	0	KGEIFLKFENMQR-TGSFKIRGAFN	174	0	YDVDNVIVPIGGGGLIAGIAVAI
THDH$YEAST	:	92	0	NTNVILKREDLLP-VFSFKLRGAYN	228	0	NKIGAVFVPVGGGGLIAGIGAYL
THRC$BACSU	threonine synthase	42	1	GIELHVKtEGVNP-TGSFKDRGMVM	175	0	EAPDVLAIPVGNAGNITAYWKGF
THRC$BRELA	:	42	0	GVQLYGKYEGANP-TGSFKDRGMVM	175	0	NAPDVLAIPVGNAGNITAYWKGF
THRC$ECOLI	:	90	2	ESDVGClElFHGP-TLAFKDFGGRF	238	3	RNQLVVSVPsGNfGDLTAGLLAk
TRP$YEAST	tryptophan synthase	367	0	GAQIWLKREDLNH-TGSHKINNALA	520	1	kLPDAVVACVGGGSNSTGMFSPF
TRPB$BACSU	tryptophan synthase β subunit	73	0	GAKIYLKREDLNH-TGSHKINNALG	225	0	TMPDKVVACVGGGSNAMGMFQAF
TRPB$BRELA	:	82	1	fARIFLKREDLVH-GGAHKTNQVIG	234	1	kLPDVVVACVGGGSNAIGMFADF
TRPB$CAUCR	:	82	0	GAKIYFKRDELNH-TGSHKINNALG	234	0	RLPDAVVACIGGGSNAIGLFHPF
TRPB$ECOLI	:	69	0	NTTLYLKREDLLH-GGAHKTNQVLG	221	0	RLPDAVIACVGGGSNAIGMFADF
TRPB$PSEAE	:	76	0	GAKIYLKREELNH-TGAHKINNCIG	228	0	RLPDSLVACIGGGSNAMGLFHPF
TRPB$PSEPU	:	78	0	GAKIFFKREELNH-TGAHKVNNCIG	230	0	RLPDSLVACVGGGSNAMGLFHEF
TRPB$SALTY	:	69	0	RTTLYLKREDLLH-GGAHKTNQVLG	221'	0	RLPDAVIACVGGGSNAIGMFADF

b) corresponding allowed amino acids:

```
AAAIAAKEDAAACIAAAFKCCAAAA     AACAACAAACGAAAAAAAAAAAF
CCCLCC FECCCF-CCSH DGCCCC     CCICCFCCCI CGGCCCGCCCCI
DGDVFF H DFGH GF I EHGGFF     DDLDFIFGNL G SDIG FFDDL
EIE GG K EHIK NG K IKMIGG     EEMEHLIIPM N  ELI GGEEM
GLI HI M FILM SH L LMNLHM     FFNGIMLLSV S  GML IIFF
KMK IL Q GKMN TI M MNQMIN     GGPIKVMMT  T  INM LLHG
MNL KM R MLNP VK W NQSVKQ     HHQLL NV      LPT MMIM
NQM LV W NMQQ  L Y QRT LS     KITMM Q       MQV TTKN
QSN M  Y QNSR  M   TT  MT     MKVNN S       NS  VVLP
RTQ T    SQTT  T   VW  NW     NL QQ T       QT  WWMQ
SVR V    TRVW  V    Y  QY     QM SR V       SV  YYQS
T S W    WS Y  W       R      RN TS W       T     RT
  T Y    YT    Y       S      SQ VT Y       V     SW
  V       V            T      TR  V               TY
          W            V      WS  W               V
          Y            W      YT  Y               W
                       Y       V                  Y
```

Fig. 1. Lysine-motif and glycine rich turn of pyridoxal-phosphate-dependent enzymes.

tryptophan synthases also the proposed pyridoxal-phosphate binding sites (7) of 2 threonine dehydratases, 2 threonine synthases and a serine dehydratase match the pattern with only a few deviations. The respective regions of the detected enzymes were used to improve our property pattern and the SWISSPROT was screened again. With this second run all threonine synthases, threonine dehydratases and serine dehydratases of the SWISSPROT were found. In addition 2 O-acetylserine sulfhydrolases were detected for which similarities of their C-terminal part to tryptophan synthase β subunit were described (19). Fig.1a shows the results of this run. Up to 5 mismatches (deviations) from the first and up to 4 from the second motif were allowed, but out of 12305 proteins in SWISSPROT only the proteins shown in fig.1a match both motifs. With a third run on the bases of property patterns derived from all regions shown in fig.1, no further sequence segment was detected before "noise" appeared.

```
                                                            4     445      4 3 4     33 3335.4. 34 345.!34.3..33  4 3 4  4 ..
a.                                                          S     TPL      L   L            K E   P  GS K R          A   G  K
THD1$ECOLI   0 --------------------------MHITYDLPVAIDDIIEAKQRLAGRIYK#GMPRSNYFSERC----KGEIFL#F#NMQR-T##F#I#GAFNKLSSLTDAEKR#GVVACS--  82
THD2$ECOLI   0 -----------------------MADSQPLSGAPEGAEYLRAVLRAPVYEAAQV###QKMEK#SSR#----DNVILV#R#DRQ#-VH#F#L#GAYAMMAGLTEEQKAHGVITAS--  86
THDH$YEAST  31 KAHLHRQHLSPSLIKLH----SELKLDELQTDNTPDYVRL---VLRSSVYDVINES#ISQGVG#SSR#----NTNVIL#R#DLL#-VF#F#L#GAYNMIAKLDDSQRNQGVIACS-- 133
THRC$BACSU   0 --------------------------MWKGLIHQYKEFLPVTDQTPALTLHEGN###IHLPK#SEQ#----GIELHV#T#GVN#-T##F#D#GMVMAVAK#KEE#NDTIM-CAS--  82
THRC$BRELA   0 --------------------------MYKGLLKQYASYLPVNEKTPDV#LMEGN###IPLLNISKQ#----GVQLYG#Y#GAN#-T##F#D#GMVMAVAK#KEE#SEAIICCAS--  83
THRC$ECOLI  28 GLFFPHDLPEFSLTEID----EMLKLDFVTRSAKILSAFIG--DEIPQEILEERVRAAFAFPAPVANV----ESDVGCLELFHG#-TLAF#DFGGRFMAQMLTHIAGD#PVTILTA- 132
sdhl$human   0 ------------------------------------------MMSGEPLHV-K##IRDSMA#SKMA----GTSVYL#MDSAQ#-S##F#I#GIGHFCKRWAKQ#CHHFV-CSS--  64
SDHL$RAT     0 ------------------------------------------MAAQE#LHV-K###RDSMA#SKVA----GTSVFL#MDSSQ#-S##F#I#GIGHLCKMKAKQ#CKHFV-CSSVV  66
SHDD$ECOLI  32 TTTLAEGLPYVGLTEQDVQDAHARLSRFAPYLAKAFPETAATGGIIESELVAIPAMQKRLEKEYQQPI----SGQLLL#KDSHL#IS##I#A#GGGIYEVL#HAEKLALEAGLLTLD 144
TRPB$SALTY   1 TTLLNPYFGEFGGMYVPQILMPALNQLEEAFVRAQKDPEFQAQFADLLKNYAGRP#A#TKCQNITAGT----RTTLYL#R#DLLH-G#AH#TNQVLGQALL#KRM#KSEIIAETG-- 110
TRPB$PSEPU   9 GPDANGLFGSFGGRYVAETLMPLVLDLAREYEAAKADPKFLEELAYFQRDYIGRPN##YFAER#TEHC---GGAKIFF#R#ELNH-T#AH#VNNCIGQVLL#KRM#KKRLIAETG-- 119
TRPB$CAUCR  12 YPDAEGRFGGFGGRYVAETLMPLVLDLGKAYADAKADPEFQAQLKSYNTHYAGRPS##YFAER#TEHF---GGAKIYF#RDELNH-T##H#INNALGQILL#MRM#KTRIIAETG-- 123
TRPB$BRELA  10 STLLPAYFGEFGGQFVAESLLPALDQLEKAFVDATNSPEFREELGGYLRDYLGRP###TECSN#PLAGEGKGFARIFL#R#DLVH-G#AH#TNQVIGQVLL#KRM#KTRIIAETG-- 123
TRPB$BACSU   4 YPNEIGRYGDFGGKFVPETLMQPLDEIQTAFKQIKDDPAFREEYYKLLKDYSGRP#A#TYADRVTEY#---GGAKIYL#R#DLNH-T##H#INNALGQALL#KKM#KT#IIAETG-- 114
TRP$YEAST    0 --KHPIRFGDFGGQYVPEALHACLRELEKGFDEAVADPTFWEDFKSLYK#Y#IGRPSS#HKAER#TEHC---QGAQIWL#R#DLNH-T##H#INNALAQVLL#KRL#KKNVIAETG-- 109
CYSK$ECOLI   0 ------------------------------------------------M#KIFEDNS#TIGHTPLVR#NRIGNGRILA#V#SRN#-SF#V#C#IGANMIWD#EKR#VL#PGVELV--  66
CYSK$SALTY   0 ------------------------------------------------M#KIYEDNS#TIGHTPLVR#NRIGNGRILA#V#SRN#-SF#V#C#IGANMIWD#EKR#VL#PGVELV--  66
                              hhhhhhhhhhhhh  hhhhhhhhhhhh      bbbbb  tttt      bbbbbbb    t t   hhhhhhhhhhh       bbbbbbb

                                                3!4.4 4.5 .455545.5.44.4       4 . 3      .3                          33335. 33 .34.
                                                 GN G A A   AAARLG K TI M      P                                          A      A
THD2$ECOLI  82 --------------------------------A##HAQGVS-LSC#M##IDGKVV#---------#KGAPK---SKVAATCD----------YSAEVVLHGDNFNDTIAKVSEIVE 143
THD1$ECOLI  86 --------------------------------A##HAQGV#-FSS####V#AL#V#---------#TATAD---IKVDAVRG----------FGGEVLLHGANFDE#KAKAIELSQ 147
THDH$YEAST 133 --------------------------------A##HAQGV#-F##KH#KIPA##V#---------#VCTPS---IKYQNVSR----------LGSQVVLYGNDFDE#KAECAKL#E 194
THRC$BACSU  82 --------------------------------T##TSA#A#-AY###ANM#CIVII---------#NGKIAF--GKLAQAVM----------YGAEIIAIDGNFDD#LKIVRSICE 144
THRC$BRELA  83 --------------------------------T##TSASA#-AY####M#CI#VI---------#EGKIAH--GKLAQAVM----------YGAEIISIEGNFDD#LKAVRNI#A 144
THRC$ECOLI 132 ------------------------------TS#DT#A#V#HAFYGLPNV#VV#LY---------#RGKISPLQEKLFCTLG----------GNIETVAIDGDFDACQALVKQAFD 198
sdhl$human  64 --------------------------------A##A#M#A#-Y##RQ##VPA##VV---------#GTTPALTIERLKNEGATCK-------VVGELLDEAFELAK#LAKNNPGWV 131
SDHL$RAT    67 QIWGSRMRGRSHGDEQPHVRSQALLPDTPSPLTA##A#M#T#-Y##R###LPA##VV---------#STTPALTIERLKNEGATVE-------VVGEMLDEAIQLAK#LEKNNPGWV 166
SDHD$ECOLI 145 DDYS---------KLLSPEFKQFFSQYSIAVGST##L#LSIG-IMS##I#F#V#VH#---------SADARAW---KKAKLRS----------HGVTVVEYEQDYGVAVEEGRKA#Q 129
TRPB$SALTY 110 --------------------------------A#QH#V#S#-L#S#L##L#CR#Y#GAKDVERQS#NVFRMRLMGAEVIPVHSGSATLK-DACNEALRDWSGSYET#HYMLGTA#G 192
TRPB$PSEPU 119 --------------------------------A#MH#V#T#-TV###F#LPCV#Y#GATDIERQQ#NVFRMKLLGAEIVPVTAGTGTLK-DAMNEALRDWVTNVEDTFYLIGTV#G 201
TRPB$CAUCR 123 --------------------------------A#QH#V#T#-TVC##F#LPCVVY#GATDVERQK#NVFRMNLLGAEVRPVSSGTGTLK-DAMNEAMRDWVTNVHDTYYLIGTA#G 205
TRPB$BRELA 123 --------------------------------A#QH#T#T#-L#C#L#M#LECVVY#GAKDVARQQ#NVYRMQLHGAKVIPVESGSGTLK-DAVNEALRDWTATFHESHYLLGTR#G 205
TRPB$BACSU 114 --------------------------------A#QH#V#P#-TV##KF#FSC#VF#GEEDVARQSLNVFRMKLLGASEVVPVTGNGTLK-DATNEAIRYWVQHCEDHFYMIGSVVG 196
TRP$YEAST  109 --------------------------------A#QH#V#T#-T#C#KF#LTC#VF#GAEDVRRQALNVFRMRILGAKVIAVTNGTKTLR-DATSEAFRFWVTNLKTTYYVVGSAIG 191
CYSK$ECOLI  66 ----------------------------EPTS##T#I#L#YV####-#Y#L#LT#------PETMSIERRKLLKALGANLV----------LTEGAK----GMKG#IQKAEEIVA 132
CYSK$SALTY  66 ----------------------------EPTN##T#I#L#YV####-#Y#L#LT#------PETMSIERRKLLKALGANLV----------LTEGAK----GMKG#IQKAEEIVA 132
                                          hhhhhhhhhhht t bbbbbb          hhhhhhtt bbbbbb      h hhhhhhhhhhht  tt bbbbb

                  4              3 34  334..3 4 .            34343..5!344.3343   .43   3 3.43   .3. 443  343.  4        33.     433 3.
                  P                P     G    A                D V   VG GG   G    K       I        DS    A    G                 A
THD2$ECOLI 144 MEG------RIFIPPYDD#KVIA#QGTIGLEIMEDL----YDV#N#IVPIGG##LIA#IAVAI#SINPTIRV#G---VQSENVHGMA#SFHS#E-----ITTHRTTGTL#DGCDVSR 242
THD1$ECOLI 148 QQG------FTWVPPFDH#MVIA#QGTL#LELLQQD----AHL#R#FVP#GG##LAA#VAVLI#QLMPQIKV#A---VEAE##ACLK#ALDA#H-----PVDLPRVGLF#EGVAVKR 246
THDH$YEAST 195 ERG------LTNIPPFDH#YVIA#QGTV#MEILRQVR-TANKIGA#FVP#GG##LIA#IGAYL#RVAPHIKT#G---VETY#AATLHNSLQRNQ-----RTPLPVVGTF#DGTSVRM 296
THRC$BACSU 145 KS#-------IALVNSVN#YRIE#QKTA#FEVCEQL---GEAP#VLAIP#GNA#NITAYWKGF#EYHEKNGTGL-PKMRGFEA-EGA#AIVRNE-----VIENPE--TI#TAIRIGN 242
THRC$BRELA 145 EE#-------ITLVNSVN#YRIE#QKTA#FEICDQL---QNAP#VLAIP#GNA#NITAYWKGFCEYEKEKGYKK-PRIHGFEA-EGA#AIVK#H-----VIEEPE--TI#TAIRIGN 242
THRC$ECOLI 199 DEE-LKVALGLNSANSINISRLLAQICYYFEAVAQLPQETRNQLV#SVPSGNF#DLTAGLLAKSLGLPVKRF#A--ATNVN#TVPRFLHDGQWS-------PKATQATLSNAMDVSQ 305
sdhl$human 132 YI#-----------PFDD#LIWE#HASIVKELKETL---WEKPGAIALS#GG##LLC#VVQGLQEVGWGDVPVI--AMETFGAHSFQ#AVKE#KLVGLRTLPKIT--SV#KALGVNT 230
SDHL$RAT   167 YIS-----------PFDD#LIWE#HTSLVKELKETL---SAKPGAIVLS#GG##LLC#VVQGLREVGWEDVP#I--AMETFGAHSFH#AVKE#KLVGLRTLPKIT--SV#KALGVNT 263
SDHD$ECOLI 130 SD#NCFFIDDENSRTLFLGYSVA#QRLK#QFAQQGRIVDADNPLF#YLPCGV##GPG#VAFGL#LAFGDHVHCF--FAEPTH#PCMLLGVHT#LH-DQISVQDINL-TA#DGLAVGR 343
TRPB$SALTY 193 PH#--------YPTIVREFQRMI#EETK#QILDKE----GRLP#A#IAC#GG#SNAI#MFADFINDTSVGLIGVEPGGHGIETGEHG#PLKH#R-VGIYFGMKAPMMQT#DGQIEES 296
TRPB$PSEPU 202 PH#--------YPAMVRDFQSII#KETR#QLQEKE----GRLP#SLVAC#GG#SNAM#LFHEFLEEPSVQIIGVEAGGHGVHTDKHA#SLNG#V-PGVLHGNRTYLLQDEDGQITDA 305
TRPB$CAUCR 206 PH#--------YPVMVRDFQSVI#AEAREQILEME----GRLP#A#VACIGG#SNAI#LFHPPFLGDEGVRIGVEAAGHGVSTDKHA#SLTG#R-PGVLHGNRTYLLQDDDGQIIDA 309
TRPB$BRELA 206 PH#--------FPTIVREFHKVISEEAK#QMLERT----GKLP#V#VAC#GG#SNAI#MFADFIDDEGVELVGAEPAGEGL##GKHG#TITN#Q-IGILHGTRSYLMRNSDGQVEES 309
TRPB$BACSU 197 PH#--------YPQVVREFQKMI#EEAKDQLKRIE----GTMP#K#VAC#GG#SNAM#MFQAFLNED-VELIGAEAAGKGI#TPLHA#TISK#T-VGVIHGSLTYLIQDEFGQIIEP 299
TRP$YEAST  192 PH#--------YPTLVRTFQSVI#KETKEQFAAMN---NGKLP#A#VAC#GG#SNST#MFSPFEHDTSVKLLGVEAGGDGV#TKFHS#TLTA#R-PGVFHGVKTYVLQDSDGQVHDT 296
CYSK$ECOLI 133 SN#----EKYLLLQQFSN#ANPEIHEKTTGPEIWED--TDGQV#VFIAG#GT##TLT#VSRYI#GTK-GKTL#S-VAVEPT##P----------------------VI#QALAGEE 218
CYSK$SALTY 133 SD#----QKYLLLQQFSN#ANPEIHEKTTGPEIWED--TDGQV#VFISG#GT##TLT#VTRYI#GTK-GKTL#T-VAVEPT##P----------------------VI#QALAGEE 218
                                 bbbbbbtthhhhhhhhhhhhhh        bbbbbbb   hhhhhhhhh     bbbbbbb

b.                                4   3 33.4 333353. 4.3 53434433  .433.535454 4    3               34.435334.4.4
                                  L        V      E    A    L E EG      E    A ALAG L                 N T V  L G G
THD2$ECOLI 243 PG-----NLTY--------EIVRELVDDI#LVSED#IRNSMIA#IQRNKVV--T#GAG#L#C#AL#SGKLDQYI-----------Q#RKT#SIIS#GNIDLSR---VSQITGFVDA- 329
THD1$ECOLI 247 IG-----DETF--------R#CQEYLDDIITVDSDAICA#MKD#F#DVRAV--A#PSG#L####MKKYIALHNI-----------RGERLAHI#S#ANVNFHGLRYVSERCELGEQR>514
THDH$YEAST 297 IG-----EETF--------RVAQQVVDEV#LVNTD#ICA#VKDIF#DTRSI--V#PSG#LSV##MKKYISTVHPE-------IDHTKN#Y#PI#S#ANMNFDRLRFVSERAVLGEGK>576
THRC$BACSU 243 PASW---DKAV--------KAAEESNGKIDEVTDD#ILH#YQLIARV##VF--A#PGSCASI##V#KQVKSGEI----------PKGSKV#AV#T#N#LKDPN-TAVDISEIKPVTL>352
THRC$BRELA 243 PASW---SYAV--------EAAEQSHGEI#LVNTD#ICA#VKDIF#DTRSI--A#PGSNAS###VIKHVESGKI----------KKGE#V#AV#T#N#LKDPD-IAISSNQLDIASV>352
THRC$ECOLI 306 PNNWPRVEELF------RRKIWQLKELGYAAVDDETTQQTMRE#-KEL#YT--S#PHA#V#YRALRDQLNPGEYGLFLGTAHPAKFKES-#EAIL#ETLDLPKELAERA-DLPLLSH>428
sdhl$human 231 VG-----AQTL--------K#FYEHPIFSEVISDQ#AVA#IEKV-DD#KIL--V#PACGA###AVYSHVIQKLQ----LEGNLRTPLPSL#VIVC#GSNISLAQLRALKEQLGMTNR>330
SDHL$RAT   264 VG-----AQTL--------K#FYEHPIFSEVISDQ#AVT#IEKFVDD#KIL--V#PACGA###AVYSGVVCRLQ----AEGRLQTPLASL#VIVC#GSNISLAQLQALKAQLGL-NE>363
SDHD$ECOLI 344 AS-----GFVG--------RAMERLLDGFYTLSDQTMYDMLGW#AQE##IR--L#PS-#L#M##PQRVCASVSYQQMHGFSAEQLR#T#HLVWAT#G#MVPEEEMNQYLAKGR---- 442
TRPB$SALTY 297 YSISAG-LDFPSVG-PQHAY#NSTGRADY#SITDD#ALE#FKT#CRH##IIPAL#SSH#L#H#LKMMREQPE-------------KEQLL#VN#S#R#DKDIFTVHDILKARGEI-- 396
TRPB$PSEPU 306 HSISAG-LDYPGIG-PEHAY#HEVKRVEY#SITDD#ALD#FHATCRL##IIPAL#SSH#L#E#IKRAPKLP--------------KDHLM#VC#S#R#DKDMQTVMNHMAAQEKQA- 405
TRPB$CAUCR 310 HSISAG-LDYPGIG-PEHSF#HDIGRAEY#STTDT#ALE#FKLCSTL##IIPAL#PAH#L#RV#EIAQELG--------------KGKIV#MN#C#R#DKDIFTVAEALGRKI---- 406
TRPB$BRELA 310 YSISAG-LDYPGVGHSTHTCTPPA-RTTL#SPTPKPSKHS-SS#ARY##IIPRTGILTRVR#RLKRAKTAEEEG-----------Q#L#ILVS#S#R#DKDVDHRAGTLEENPELIL>416
TRPB$BACSU 300 YSISAG-LDYPGIG-PEHAY#HKSGRVTYDSITDE#AVD#LKL#S#K##ILPAI#SAH#L#K#FK#AKGMD-------------RGQLILVC#S#R#DKDVNTLMNVLEEEVKRHV 400
TRP$YEAST  249 HSVSAG-LDYPGVG-PELAY#KSTGRAQFIAATDAQALLGFKL#SQL##IIPAL#SSH#VYG#CE#AKTMK--------------PDQHL#INIS#R#DKDVQSVAEVLPKLGPKIG>407
CYSK$ECOLI 249 IKPGPHKIQGIGAG-FIPAN#DLKLVDKVIGITNE#AISTARR#M#E##ILAGISSGA#V#A#LK#QEDES-------------FT#KNI#VI#PSS#ERYLSTALFADLFTEKELQ 323
CYSK$SALTY 167 IKPGPHKIQGIGAG-FIPGN#DLKLIDKV#GITNE#AISTARR#M#E#VFLAGISSGA#V#AALK#QEDES-------------FT#KNI#VI#PSS#ERYLSTALFADLFTEKELQ 323
                                 hhhhhhhhh  bbbbbb hhhhhhhhhhhhh        hhhhhhhhhhhhhhhhh            bbbbbbb      hhhhhhhhhhhhhh
```

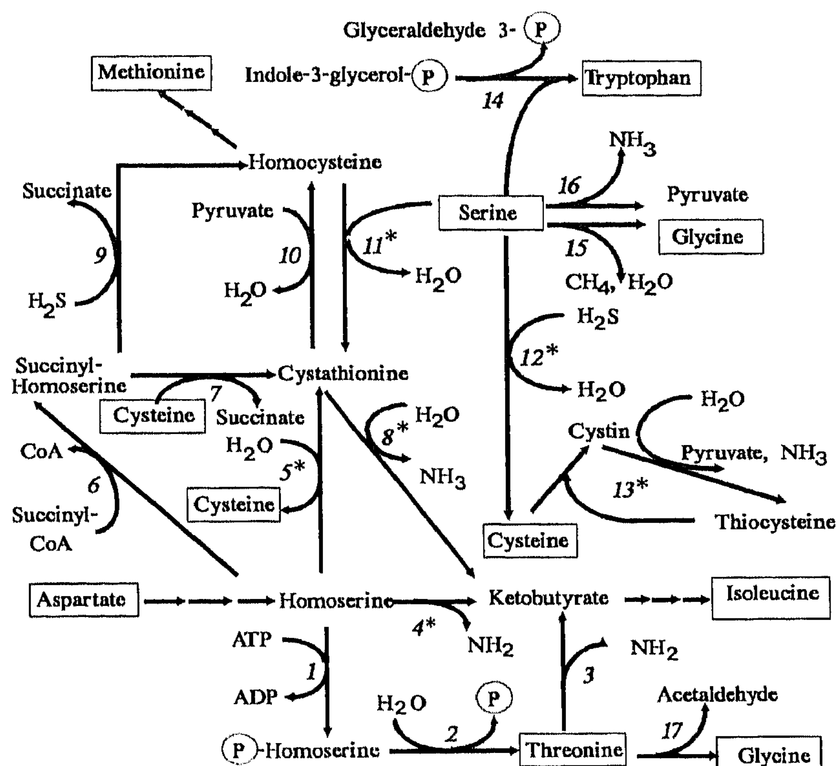

Fig. 3. Metabolic map of amino acids and pyridoxal-phosphate-dependent enzymes.

To verify the indicated similarities among the detected enzymes we have used our multiple alignment program PULIN. The similar regions could be extended and we propose now an alignment, wich reveals a similar folding topology in long segments of the included enzymes (fig.2).

The similar sequence segments agree with structural findings, because deletions and insertions mostly occur in regions between corresponding secondary structure elements of tryptophan synthase (fig.2). Our alignment with O-acetylserine sulfhydrolase is somewhat different to that proposed by Levy and Danchin (19). So we have related the potential pyridoxal-phosphate-binding site in O-acetylserine sulfhydrolase to a region in tryptophan synthase around the lysine-87 (numbering according to the enzyme from salmonella typhimurium, fig.2) as indicated by our property patterns (fig.1).

Only in 3 positions amino acids were found to be absolutely conserved in all sequences (marked by an exclamation mark in fig.2). One of them is the essential lysine covalently bound to pyridoxal-phosphate, the second is a glycine of the second motif used for the pattern construction and the third is located in a

---

Fig. 2. Conserved sequences among pyridoxal-phosphate-dependent enzymes.

region surprisingly found to be the most conserved one corresponding to residues 112 to 134 in tryptophan synthase from salmonella thyphimurium (fig.2). This region comprises an α-helix with a subsequent turn and consists of a cluster of tiny amino acids. Further conserved regions comprise the motifs used for pattern construction. All three most conserved regions are not located in the β-sheet core so that they seem to have essential functions in pyridoxal-phosphate-binding.

Discussion

The pairwise identities between enzymes of the different subgroups in fig.2 range between 15% and 19%. These values are typical for the so called 'twilight zone' in which usually automatic alignment procedures fail (20,21), but background information argues in favour of their similarities. All enzymes involved in our alignment and proposed to be homologous have some features in common: (i) they all need pyridoxal-phosphate as a cofactor, (ii) there are some similarities in their reaction mechanisms via an intermediate and (iii) they all catalyse successive steps in biochemical pathways (fig.3). The third point led to the hypothesis of an evolution of these pathways from an ancestor pyridoxal-phosphate-binding protein. In fig.3 some successive reactions of the aspartate family and their connections to tryptophan synthase are shown. All of them are catalysed by pyridoxal-phosphate-dependent enzymes. Therefore we have also investigated other enzymes of this scheme (fig.3) not found by our pattern and alignment procedures. Homologies between cystathionine γ-synthase (EC 4.2.99.9) and cystathionine β-lyase (4.4.1.8) were detected (22) and we have also found a high level of similarity to O-acetylhomoserine thiolyase (EC 4.2.99.10). O-acetylhomoserine thiolyase can catalyse the same reaction as O-acetylserine sulfhydrolase which is, in turn, homologous to tryptophan synthase. Despite an amino acid identity of 11% (10 gaps were introduced by PULIN, data not shown) a homology could not be verified. The identical positions are in contrast to the alignment of O-acetylhomoserine thiolyase with cystathionine β-lyase and cystathionine γ-synthase.

Nevertheless, the evolution of these pathways can be explained as a series of gene duplication events. In various aminotransferases or decarboxylases only the region responsible for the recognition of the respective amino acid had to be changed

after gene duplication to explain the different positions in the pathways. In contrast, the family of similar pyridoxal-phosphate-dependent enzymes discussed here reveals very early gene duplication events leading to the development of biochemical pathways.

The suggested homology between tryptophan synthase (β subunit) and the other enzymes presented here can be used to build topology models of all these proteins. The conserved regions and positions show that despite the expected variability of pyridoxal-phosphate-dependent enzymes (see introduction) there seems to exist only a limited set of protein topologies compatible with pyridoxal-phosphate-binding via lysine.

References

1. Sprang, S.R., Acharya, K.R., Goldsmith, E.J., Stuart, D.I., Varvill, K., Fletterick, R.J., Madsen, N.B. and Johnson, L.N. (1988) Nature 336, 215-221.
2. Barford, D. and Johnson, C.N. (1989) Nature 340, 609-616.
3. Ford, G.C., Eichele, G. and Jansonius, J.N. (1980) Proc. Natl.Acad.Sci.USA 77, 2559-2563.
4. Arnone, A., Rogers, P.H., Hyde, C.C., Briley, P.D., Metzler, C.M. and Metzler D.E. (in Transaminases (Christen, P. and Metzler, D.E. eds.) pp138-155, John Wiley and Sons, New York
5. Harutyunyan, E.G., Malashkevich, V.N., Tersyan, S.S., Kochkina, V.M., Torchinski Y.M. and Braunstein, A.E. (1982) FEBS Lett. 138, 113-116.
6. Hyde, C.C., Ahmed, S.A., Padlan, E.A., Miles E.W. and Davies, D.R. (1988) J.Biol.Chem. 263, 17857-17871.
7. Rossmann, M.G., Lilias, A., Bränden, C.I. and Banaszak, L.I. (1975) in The enzymes (Boyer,P.D.ed.) Vol. 11, 3nd edition, 61-101, Academic Press, New York
8. Parsot, C. (1986) EMBO J. 5, 3013-3019.
9. Mehta, K., Hale, T.I.and Christen, P. (1989) Eur.J.Biochem. 186, 249-253.
10. Bork, P. (1989) FEBS Lett. 257, 191-195.
11. Bork, P. and Grunwald, C. (1990) Eur.J.Biochem., in press
12. Lipman, D.J. and Pearson, W.R. (1985) Science 227, 1435-1441.
13. Santibanez, M. and Rohde, K. (1987) CABIOS 3, 111-135.
14. Bork, P. and Grunwald, C. (1989) Stud.Biophys. 129, 231-240.
15. Gribskov, M., McLachlan, A.D. and Eisenberg, D. (1987) Proc. Natl.Acad.Sci.USA 84, 4355-4358.
16. Barton, G.I. and Sternberg, M.J.E. (1987) J.Mol.Biol. 198, 327-337.
17. Vingron, M. and Argos, P. (1989) CABIOS 5, 115-121.
18. Dayhoff, M. (1978) Atlas of Protein sequence and structure (National Biomedical Research Foundation), Vol.5, Suppl. 3
19. Levy, S. and Danchin, A. (1988) Mol.Microbiol. 2, 777-783.
20. Doolittle, R.F. (1985) Sci. Amer. 253, 74-83.
21. Taylor, W.R. (1988) Prot.Eng. 2, 77-86.
22. Belfaiza, J., Parsot, C., Martel, A., La Tour, C.B., Margarita, D., Cohen, G.N. and Saint-Girons, I. (1986) Proc. Natl.Acad.Sci.USA 83, 867-871.
23. Ogawa, H., Gomi, T., Kanishi, K., Date, T., Nakashima, H., Nose, K., Matsuda, Y., Peruino, C., Pitot, H.C. and Fujioka, M. (1989) J.Biol.Chem. 264, 15818-158123.