The protein kinases of budding yeast: six score and more

Tony Hunter and Gregory D. Plowman

The completion of the budding yeast genome sequencing project has made it possible to determine not only the total number of genes, but also the exact number of genes of a particular type[1-3]. As a consequence, we now know exactly how many protein kinases are encoded by the yeast genome, a number of considerable interest because of the importance of protein phosphorylation in the control of so many cellular processes.

BUDDING YEAST has 113 conventional protein kinase genes, corresponding to ~2% of the total genes (see Table 1 in centrefold). More than 60% of these protein kinases have either known or suspected functions; the remainder are novel, and functional analysis awaits. In terms of defined functions encoded by the yeast genome, protein kinases come a close second behind transcription factors[3].

What can be learnt from knowing all the protein kinases encoded by a single eukaryotic genome? One obvious outcome is that it is possible to say whether a protein kinase identified and characterized in another organism has a homologue in budding yeast. Such a homologue might either have a known function in yeast, or its function can be tested by genetic studies. Equally important is the recognition of protein kinase subfamilies present in higher eukaryotes that are absent from yeast. Whereas all eukaryotes have similar requirements for DNA replication, transcription, translation and energy metabolism, it is reasonable to expect that there might be many protein kinases unique to multicellular organisms that function in cellular communication, both between cells, tissues and the environment, such as protein-tyrosine kinases. Conversely, there might be protein kinases unique to budding yeast.

T. Hunter is at The Salk Institute,
10010 North Torrey Pines Road, La Jolla,
CA 92037, USA.
Email: hunter@salk.edu
G. D. Plowman is at SUGEN, Inc.,
515 Galveston Drive, Redwood City,
CA 94063, USA.
Email: plowman@sugen.sf.ca.us

Multiple alignment and parsimony analysis of catalytic domain sequences (Fig. 1) categorizes the yeast protein kinases into subfamilies based on structural relatedness. From such a classification, one can infer functional similarities, including regulation of catalytic activity, substrate specificity and cellular localization. This information is of particular value for understanding the function of the numerous uncharacterized yeast open reading frames that exhibit protein kinase motifs. One can also determine which protein kinases subfamilies are conserved or expanded in other organisms and which are unique to yeast.

MAP kinase pathways

Pseudohyphal development. One of the virtues of knowing all yeast protein kinases is that it delimits the number of protein kinases in a particular subfamily. For example, five members of the mitogen-activated protein (MAP) kinase family, Fus3, Kss1, Hog1, Mpk1/Slt2 and Smk1, had been identified and functionally characterized before the completion of the genome project[4]. However, both haploid invasive growth and diploid pseudohyphal development of budding yeast are known to require Ste11 and Ste7, which are a MAP kinase kinase kinase (MAPKKK) and a MAP kinase kinase (MAPKK), respectively. These kinases normally function in the mating pheromone response pathway to activate the Fus3 (or Kss1) MAP kinase, and also Ste12, a transcription factor that is phosphorylated and activated by Fus3 (Ref. 5). This suggested that a MAP kinase would be required for pseudohyphal development, but none of the five characterized MAP kinases proved to be essential for haploid invasive

growth, either singly or in combination (H. Madhani and G. Fink, pers. commun.).

The putative protein kinase YKL161C clusters with the MAP kinases (Fig. 1) and might be another MAP kinase (although it has a KxY rather than TxY motif in the activation loop), but similarly, this gene is not essential for pseudohyphal development. As Ste7 normally phosphorylates the Thr and Tyr in the TxY motif in the activation loop of Fus3 and Kss1, one possibility is that Ste7 might phosphorylate and activate another protein kinase with a related activation-loop motif. Three other protein kinases, Kin3/Npk1, Ssn3/Srb10 and Ime2, have a TxY activation-loop sequence and might, in principle, be regulated by Ste7 phosphorylation. Ssn3/Srb10 is a cyclin-dependent kinase (CDK) and is unlikely to be a Ste7 target. Ime2, however, has motifs characteristic of proline-directed protein kinases, like the MAP kinases, and is on the same major branch as the MAP kinases (Fig. 1), making it a potential candidate, although Ime2 is normally only produced during meiosis[6].

With regard to the protein kinases lying upstream of the MAP kinases, there are four MAPKKs (Ste7, Mkk1, Mkk2 and Pbs2: STE7/MEK family), and four MAPKKKs (Ste11, Bck1, Ssk2 and Ssk22: STE11/MEKK family). These have been assigned to the Fus3+Kss1, Mpk1 or Hog1 MAP kinase pathways by genetical and biochemical analyses[4]. This leaves both the Smk1 MAP kinase, which is required for spore wall assembly[7], and YKL161C without their own specific MAPKK and MAPKKK. Possibly, a MAPKKK/MAPKK combination used for one of the other MAP kinase pathways is also used for these predicted MAP kinases. Interestingly, however, the Smk1 pathway might have a specific MAPKKK kinase, as the Sps1 protein kinase (NRK/MESS family), which lies upstream of Smk1 (Ref. 8), is related to Ste20, a MAPKKK kinase in the Fus3 MAP kinase pathway.

The Ste20 family. Ste20 is a Cdc42-activated protein kinase[9] that is required in the pheromone response MAP kinase pathway upstream of Ste11 (Refs 10, 11). A *ste20* mutant is viable and only shows a defect in response to mating pheromone[11]. Cla4 is a second member of the Ste20 family, which also interacts with Cdc42, and has been implicated in polarized cell growth, budding and cytokinesis, but a *cla4* mutant is viable despite a cytokinesis defect[12].

A double *ste20/cla4* mutant, however, cannot undergo cytokinesis, implying

Table I. Classification of *Saccharomyces cerevisiae* protein kinases[a]

AGC group (17 members)

PKA family: cAMP-dependent (Similar to PKA_h)
 TPK1/PKA1/SRA3/PK25/J0541/YJL164C PKA 1, catalytic subunit
 TPK2/PKA2/YKR1/P1855/YPL203W PKA 2, catalytic subunit
 TPK3/PKA3/YKL630/YKL166C PKA 3, catalytic subunit

PKC family: DAB-activated, PL-dependent (Similar to PKC_h)
 PKC1/STT1/HP02/CLY15/YBL0807/YBL105C PKC; regulates MAP kinase cascade involved in regulating cell wall metabolism

AGC family (Similar to SCK1_sp, RACa_h, AKT_h – no diacylglycerol-binding domain)
 SCH9/KOM1/YHR205W Ser/Thr protein kinase activated by cAMP; overproduction can suppress *cdc25* mutant
 YPK1/YKL126W Ser/Thr protein kinase with similarity to PKC
 YPK2/YKR2/YM9718.03/YMR104C Ser/Thr protein kinase with similarity to Ypk1p

S6K (70 kDa) family (Similar to KAD5_sp, SGK_r)
 KIN82/YCR1153/YCR091W Ser/Thr protein kinase of unknown function
 YNR047W/N3449 Ser/Thr protein kinase of unknown function

DBF2 family (Similar to KAIB_sp, NDR_h)
 DBF2/G4643/YGR092W Ser/Thr protein kinase similar to Dbf20p; required for anaphase/telophase
 DBF20/P8283.6/YPR111W Cell-cycle protein kinase similar to Dbf2p; involved in M-phase termination

PKA-related family (Unique to *S. cerevisiae*)
 YOL100W/HRC1081/O0784 Ser/Thr protein kinase of unknown function
 YDR490C/D8035.33 Ser/Thr protein kinase of unknown function
 YDR466W/D8035.10 Ser/Thr protein kinase of unknown function

Other AGC family (Similar to CEK1_sp, COT1_nc, NDR_h, LATS_dm, MAST205_m)
 YFL033C Ser/Thr protein kinase with similarity to *Schizosaccharomyces pombe* CEK1 protein kinase
 YNL161W/N1727 Putative Ser/Thr protein kinase of unknown function
 *YBR028C/YBR0312 Ser/Thr protein kinase with similarity to Ypk2p/Ykr2p and Ypk1p

CaMK group (16 members)

CaMK family: Ca^{2+}-calmodulin regulated (Similar to CaMKI_h)
 CMK1/YFR014C Ca^{2+}-calmodulin-dependent Ser/Thr protein kinase (CaM kinase), type I
 CMK2/O2325/YOL016C CaM kinase type II
 RCK1/G1854/YGL158W Ser/Thr protein kinase with similarity to Cmk1p, Cmk2p and Cmk3p. (Sequence updated)
 RCK2/CMK3/CLK1/L9672.6/YLR248W CaM kinase

SNF1/AMPK family (Similar to AMPK_h, NPK5_nt, PAR1_ce)
 SNF1/CAT1/CCR1/PAS14/HAF3/D8035.20/YDR477W Ser/Thr protein kinase essential for derepression of glucose-repressed genes; acts with Snf4p
 KIN1/YD9727.17/YDR122W Ser/Thr protein kinase; similar to Kin2p and *S. pombe* KIN1
 KIN2/L8004.3/L2546/YLR096W Ser/Thr protein kinase; similar to Kin1p and *S. pombe* KIN1
 KIN4/KIN31/(KIN3)/O5220/YOR233W Ser/Thr protein kinase; similar to Kin1p and Kin2p; catalytic domain is most similar to Snf1p
 YPL141C/LPI5 Ser/Thr protein kinase with similarity to Kin4p
 YPL150W/P2597 Ser/Thr protein kinase of unknown function. (Not in YPD listing)

GIN4 family (Similar to SNF1_sc)
 GIN4/D9719.13/YDR507C Ser/Thr protein kinase with similarity to Yc1024p; growth inhibitory protein
 YCL024W Protein with similarity to Snf1p
 HSL1/YKL453/YKL101W Ser/Thr protein kinase that interacts genetically with histone mutations

Other CaMK family (Similar to Z71478_sp, MLCK_dd, CaMKI_h)
 MEK1/MRE4/O6357/YOR351C Ser/Thr protein kinase required for meiotic recombination; contains forkhead-associated (FHA) nuclear signalling domain
 DUN1/ORF2370/YDL101C Protein kinase necessary for induction of Rnr3p and DNA repair genes after DNA damage; contains FHA domain
 *YMR291W/ORF530348 Ser/Thr protein kinase of unknown function

CMGC group (21 members)

CDK family (Similar to CDK2_h)
 CDC28/CDK1/SRM5/HSL5/YBR1211/YBR160W Cyclin-dependent protein kinase (CDK) essential for completion of START and for mitosis; associates with Cks1p and cyclins. PSTAIRE in kinase subdomain III
 PHO85/P7102_18A/YPL031C CDK that interacts with cyclin Pho80p to regulate phosphate pathway. PSTAIRE in kinase subdomain III

Table I. Classification of *Saccharomyces cerevisiae* protein kinases (contd)

CDK family contd

CAK1/CIV1/YFL029C CDK-activating kinase (Ser/Thr protein kinase) responsible for *in vivo* activation of Cdc28p. PHNAKFE in kinase subdomain III

SSN3/UME5/SRB10/(ARE1)/P7102_08/YPL042C Ser/Thr CDK of the RNA polymerase II holoenzyme complex and mediator (SRB) subcomplex. SQSACRE in kinase subdomain III. TLY in kinase subomain VIII

KIN28/ORF2330/YDL108W Ser/Thr CDK component of transcription initiation factor TFIIH; phosphorylates carboxy-terminal domain (CTD) of RNA polymerase large subunit. DMSAIRE in kinase subdomain III

MAPK family (Similar to SPK1_sp, ERK_h)

KSS1/G4149/YGR040W Ser/Thr protein kinase; redundant with Fus3p for induction of mating-specific genes by mating pheromone. TEY in kinase subdomain VIII (activation loop)

FUS3/DAC2/YBL0303/YBL03.21/YBL016W Ser/Thr protein kinase required for cell-cycle arrest and for cell fusion during mating. TEY in activation loop

HOG1/SSK3/L9354.2/L2931/YLR113W Ser/Thr protein kinase; involved in high-osmolarity signal transduction pathway. TGY in activation loop

SLT2/MPK1/SLK2/BYC2/YHR030C Ser/Thr protein kinase involved in the cell wall integrity pathway. TEY in activation loop

YKL161C/YKL615 Ser/Thr protein kinase of unknown function. KGY in activation loop

SMK1/YP9499.10/YPR048W Sporulation-specific MAP kinase required for completion of sporulation. TNY in activation loop

GSK3 family (Similar to SHAGGY_dm, GSK3_h)

MCK1/(YPK1)/N0392/YNL307C Ser/Thr/Tyr protein kinase (meiosis and centromere regulatory kinase); positive regulator of meiosis and spore formation

YOL128C/O0530/ORF1209713 Ser/Thr protein kinase of unknown function

MDS1/RIM11/GSK3/YM9375.08/YMR139W Ser/Thr protein kinase; homologue of mammalian GSK3

MRK1/D2459/D2461/YDL079C Ser/Thr protein kinase with similarity to Mds1p

CLK family

KNS1/L1224/YLL019C Ser/Thr protein kinase of unknown function. (Similar to CLK_h)

YAK1/YJL141C Ser/Thr protein kinase that suppresses loss of Tpk1p + Tpk2p + Tpk3p. (Similar to KA23_sp, MNB_dm, MNB_h)

YMR216C/YM8261.10 Putative Ser/Thr protein kinase; has similarity to Cdc31p. (Similar to DSK1_sp, SRPK1_ce, U52111_h). (Not in YPD listing)

IME2/SME1/J0817/YJL106W Ser/Thr protein kinase and positive regulator of sporulation genes essential for initiation of meiosis. TAY in activation loop. (Similar to MAK_r, p34_h, CDK2_h)

Other CMGC family (Similar to PITSLRE_h, CHED_h)

SGV1/BUR1/P9584.8/YPR161C Ser/Thr protein kinase involved in pheromone adaptation pathway and in cell cycle. PITAQRE in kinase subdomain III

CTK1/YKL139W CTD kinase a subunit; CDK that phosphorylates CTD of RNA polymerase II large subunit. PITSIRE in kinase subdomain III

STE11/STE20 group (10 members)

STE11/MEKK family (Similar to BYR2_sp, NPK1_nt, MEKK_h)

STE11/L8039.10/YLR362W Ser/Thr protein kinase; component of the pheromone pathway and a pathway regulating pseudohyphal development

BCK1/(SLK1)/SSP31/LAS3/SAP3/J0906/YJL095W Ser/Thr protein kinase; involved in the cell wall integrity pathway

SSK2/N3276/YNR031C MAP kinase kinase kinase (MEKK) of the high osmolarity signal transduction pathway

SSK22/YCR073C MEKK with strong similarity to Ssk2p; participates in the high osmolarity signal transduction pathway

STE20/PAK family (Similar to PAK_dm, PAK1_h, PAK65_h, RAC_h)

STE20/YHL007C Ser/Thr protein kinase in the pheromone pathway; also participates in pathway regulating pseudohyphal development

CLA4/ERC10/N0450/YNL0450/YNL298W Ser/Thr protein kinase required for cytokinesis; has similarity to Ste20p

YOL113W/HRA655/O0722 Ser/Thr protein kinase with similarity to Ste20p

NRK/MESS family (Similar to MESS1_m, ZC504.4_ce)

NRK1/KIC1/H8263.14/YHR102W Ser/Thr protein kinase that interacts with Cdc31p

SPS1/D9719.27/YDR523C Ser/Thr protein kinase involved in middle/late stage of meiosis

***CDC15/YAR019C** Protein kinase of the MAP kinase kinase kinase family essential for late nuclear division. (Similar to MESS1_m, CDC7_sp, MST1_h)

STE7/MEK group (8 members)

STE7 family (Similar to HST7_ce, MEK1_h)

STE7/D1525/YDL159W Ser/Thr/Tyr protein kinase of MAP kinase kinase (MEK) family; component of the pheromone pathway and a pathway regulating pseudohyphal development

PBS2/HOG4/SFS4/SSK4/OSR1/J0699/YJL128C Ser/Thr/Tyr protein kinase of the MEK kinase family; essential component of the high-osmolarity signal transduction pathway

MKK1/SSP32/O5095/YOR231W Ser/Thr/Tyr protein kinase of the MEK family involved in cell wall integrity pathway

MKK2/SSP33/LPI6/YPL140C Ser/Thr/Tyr protein kinase of the MEK family involved in cell wall integrity pathway. (Sequence updated)

Table I. Classification of *Saccharomyces cerevisiae* protein kinases (contd)

NIMA/NEK family (Similar to NIMA_en, NIMA1_h, NEK1_h)
 KIN3/NPK1/FUN52/YAR018C Ser/Thr protein kinase; null mutation has no phenotype. TTY in kinase subdomain VIII (activation loop)

NEK-like family (Similar to F35G12.3_ce, weakly to NEK1_h)
 YNL020C/N2823 Ser/Thr protein kinase of unknown function
 YIL095W/(PAK1) Ser/Thr protein kinase of unknown function
 YBR059C/YBR0419 Ser/Thr protein kinase of unknown function

Other group (24 members)

Casein kinase I family (Similar to CKI_h)
 YCK1/CKI2/YHR135C Casein kinase I (CKI) isoform
 YCK2/(CKI1)/(CKI)/N1755/YNL154C CKI isoform
 YCK3/CKI3/YER123W CKI isoform
 HRR25/P1850/YPL204W CKI, Ser/Thr/Tyr protein kinase; associated with DNA repair and meiosis

Casein kinase II family (Similar to CKA2_h)
 CKA1/YIL035C CKII, catalytic (α) subunit
 CKA2/O2810/YOR061W CKII, catalytic (α) subunit
 *CDC7/SAS1/OAF2/D2855/YDL017W Protein kinase required for initiation of DNA synthesis, for commitment to sporulation, for DNA repair
 and for meiotic recombination. (Similar to HSK1_sp)

NPR/HAL5 family (Unique to *S. cerevisiae*)
 HAL5/J0531/YJL165C Ser/Thr protein kinase involved in salt and pH tolerance
 YKL168C/YKL632 Ser/Thr protein kinase of unknown function
 SAT4/YCR101/YCR046/YCR008W Protein with similarity to Npr1p protein kinase
 YJR059W/J1725 Putative Ser/Thr protein kinase of unknown function
 PTK1/YKL198C Ser/Thr protein kinase that enhances spermine uptake. (Frame shift corrected)
 NPR1/N1631/YNL183C Ser/Thr protein kinase involved in regulating transport systems for nitrogen nutrients under conditions of nitrogen
 catabolite derepression
 YDL214C/D1014 Ser/Thr protein kinase with similarity to Npr1
 YDL025C/D2810 Protein with similarity to protein kinase Npr1p
 YOR267C/O5420 Ser/Thr kinase protein kinase with similarity to Npr1p

ELM family (Unique to yeast, similar to D45882_sp)
 PAK1/SYGP-ORF45/YER129W Protein kinase capable of suppressing DNA polymerase α mutations
 YGL179C/BIE560/G1618 Ser/Thr protein kinase with similarity to Elm1p and Kin82p
 ELM1/YKL261/YKL048C Ser/Thr protein kinase regulating pseudohyphal development

RAN family (Similar to RAN_sp, p78_h)
 SHA3/SKS1/LPB5/YPL026C Ser/Thr protein kinase; suppressor of *hta1* mutations that cause aberrant transcription
 YDR247W/YD8419.14 Ser/Thr protein kinase with similarity to *S. pombe* RAN1 negative regulator of sexual conjugation and meiosis
 *KSP1/YHR082C Ser/Thr kinase with similarity to CKII

PIM-like family (Similar to PIM2_m, KIAA0135_h)
 *YAL017W/YAL002/FUN31 Ser/Thr protein kinase of unknown function
 *YOL045W/O2034/YOL044W Ser/Thr protein kinase of unknown function

Unique kinases (17 members) (No similar *S. cerevisiae* kinases)

Kinases with possible homologues in other species
 *CDC5/PKX2/MSD2/YM8270.03/YMR001C Ser/Thr protein kinase required for exit from mitosis; ts mutants block after nuclear division. (Similar
 to PLK-1_h, POLO_dm)
 *IPL1/P1820/YPL209C Ser/Thr protein kinase involved in chromosome segregation. (Similar to AUR_dm)
 *IRE1/ERN1/YHR079C Protein kinase and type I membrane protein involved in signal transduction from ER lumen to nucleus; part of the unfolded
 protein response. (Similar to C41C4.4_ce)
 *VPS15/VPT15/(VPL19)/YBR0825/YBR097W Ser/Thr protein kinase involved in vacuolar protein sorting. (Similar to ZK930.1_ce)
 *YPL236C/P1057 Protein of unknown function. No GxG (Similar to C3H1.13_sp) (Not in YPD)
 *YGL180W/G1615 Ser/Thr protein kinase of unknown function. (Similar to UNC-51_ce, PLO1_sp)
 *SWE1/J0406/YJL187C Ser/Tyr dual-specificity protein kinase able to phosphorylate Cdc28p on tyrosine and inhibit its activity. (Similar to Wee1_sp
 and MLK_h)
 *SPK1/RAD53/MEC2/SAD1/P2588/YPL153C Ser/Thr/Tyr protein kinase with a checkpoint function in S and G2. Contains FHA domain. (Similar
 to CDS1_sp)
 *MPS1/RPK1/D2785/YDL028C Ser/Thr/Tyr protein kinase involved in spindle pole body duplication (Similar to ESK_m, TTK_h)
 *YKL116C/YKL516 Ser/Thr protein kinase with similarity to *S. pombe* NIM1 protein kinase. (33% identity to p78_h)

Table I. Classification of *Saccharomyces cerevisiae* protein kinases (contd)

Kinases with possible homologues in other species contd

*GCN2/AAS1/D9954.16/YDR283C Ser/Thr protein kinase that regulates initiation of translation by phosphorylation of eIF2α (Sui2p) (Similar to EIF2aK_r, HRI_r)

*YBR274W/YBR1742 Protein kinase with similarity to members of the growth factor and cytokine receptor family. (Similar to CHK1-sp, CHK1_ce, SNF1_r)

*YGR262C/G9334 Protein with similarity to apple tree CaM-binding protein kinase PIR:JQ2251. Lacks GxG – not in alignment. (Similar to O–sialoglycoprotein endopeptidase from *Methanococcus jannaschii*)

Kinases without known homologue

*BUB1/G7542/YGR188C Ser/Thr protein kinase and checkpoint protein required for cell-cycle arrest in response to loss of microtubule function. (Amino terminus similar to MAD3_sc)

*YKL171W/YKL635 Ser/Thr protein kinase of unknown function

*YGR052W/G4329 Protein of unknown function

*YPR106W/P8283.9 Protein with similarity to protein kinases Gcn2p; galactosyltransferase-associated protein kinase P58/Gtap, and the Raf proto-oncoprotein

Atypical protein kinases (1 member)

*YGR080W/G4583 Protein with similarity to human tyrosine kinase A6 PIR:A55922.

Miscellaneous kinases

Phosphatidylinositol-kinases

TOR1/DRR1/J1803/YJR066W Phosphatidylinositol kinase (PI kinase) homologue involved in cell growth and sensitivity to the immunosuppressant rapamycin

TOR2/DRR2/YKL203C PI kinase homologue involved in cell growth and sensitivity to the immunosuppressant rapamycin, similar to Tor1p

VPS34/VPT29/(VPL7)/END12/L9672.10/YKR240W PI 3-kinase required for vacuolar protein sorting; activated by protein kinase Vps15p

PIK1/PIK41/PIK120/N0795/YNL267W PI 4-kinase; generates PtdIns(4)P

STT4/L2142.4/YLR305C PI 4-kinase; mutants are staurosporine-sensitive and suppressible by overproduction of Pkc1p

MEC1/ESR1/SAD3/YBR1012/YBR136W Checkpoint protein required for mitotic growth; DNA repair and mitotic recombination (PI kinase homologue)

TEL1/YBL0706/YBL088C Protein involved in controlling telomere length; might have PI 3-kinase or protein kinase activity

YHR099W Protein with weak similarity to Tor1p and Tor2p; possible PI kinase homologue

FAB1/YFR019W Probable PtdIns(4)P 5-kinase involved in orientation or separation of mitotic chromosomes

MSS4/YD8142A.05/YD8142.05 Potential PtdIns(4)P 5-kinase; multicopy suppressor of *stt4* mutation

**Guanylate kinases**

GUK1/D9461.39/YDR454C Guanylate kinase

Histidine protein kinases

SLN1/YPD2/YIL147C Two-component signal transducer with both a His kinase domain and a receiver domain that functions in the high osmolarity signal transduction pathway

YIL042C Related to the mitochondrial branched-chain α-ketoacid (BKCD) and pyruvate dehydrogenase (PDH) kinases, which are protein-serine kinases

Misclassified kinases

MSS2/D2340/YDL107W/ORF2340 Involved in the expression of mitochondrial cytochrome C oxidase subunit 2 (COX2). (Ser/Thr protein kinase that suppresses the growth defect of *snf3* mutants on low glucose.)

YDR109C/YD9727.05 Protein with similarity to FGGY protein kinase family

YJR061W Putative Ser/Thr protein kinase of unknown function. (Similar to YKL200c_sc, YKL201c_sc)

YLR063W/L2174 Ser/Thr protein kinase of unknown function (Unique)

YML059C/YM9958.03 Protein kinase of unknown function (Similar to ZK370.4_ce, M110.7_ce)

YMR192W/YM9646.04 Ser/Thr protein kinase of unknown function. (Leucine repeat and possible coiled coil. Similar to YPL249C_sc, U49940_ce)

YGL059W/G3441 Protein with similarity to rat branched-chain α-ketoacid dehydrogenase kinase PIR:U27456

YGL227W/G0958 Protein with similarity to *Dictyostelium discoideum* non-receptor tyrosine kinase U32174. 37% identity over 64 residues in amino-terminal non-catalytic domain

YOR287C/O5492 Protein with weak similarity to PITSLRE protein kinase isoforms

aThe budding yeast protein kinases are subdivided into distinct families based on structural similarity in their catalytic domains. The overall classification is based on that devised by Hanks and Hunter[38]. Individual kinases are listed by their preferred gene name as established by the Saccharomyces Genome Database at Stanford, followed by their synonyms and a brief description as maintained in the Yeast Protein Database (YPD) available on the Internet at http://www.proteome.com. Additional notes, sequence corrections or close homologues are in parentheses following each entry. Protein kinases preceded by an asterisk share only weak similarity to other members. Entries listed as misclassified kinases are flagged as protein kinases in YPD, but have no structural similarity to the protein kinase family and were excluded from this analysis (some of these represent non-catalytic regulatory domains). Three additional open reading frames were identified that encode protein kinases that were not present in the YPD listing: YPL150W (Group IIB), YMR216C (Group IIID) and YPL236C (Group VII). These new protein kinases were recognized following a comprehensive analysis of the complete yeast DNA sequence using the MPSRCH software (Oxford Molecular) implementation of the Smith Waterman algorithm on a Maspar parallel computer.
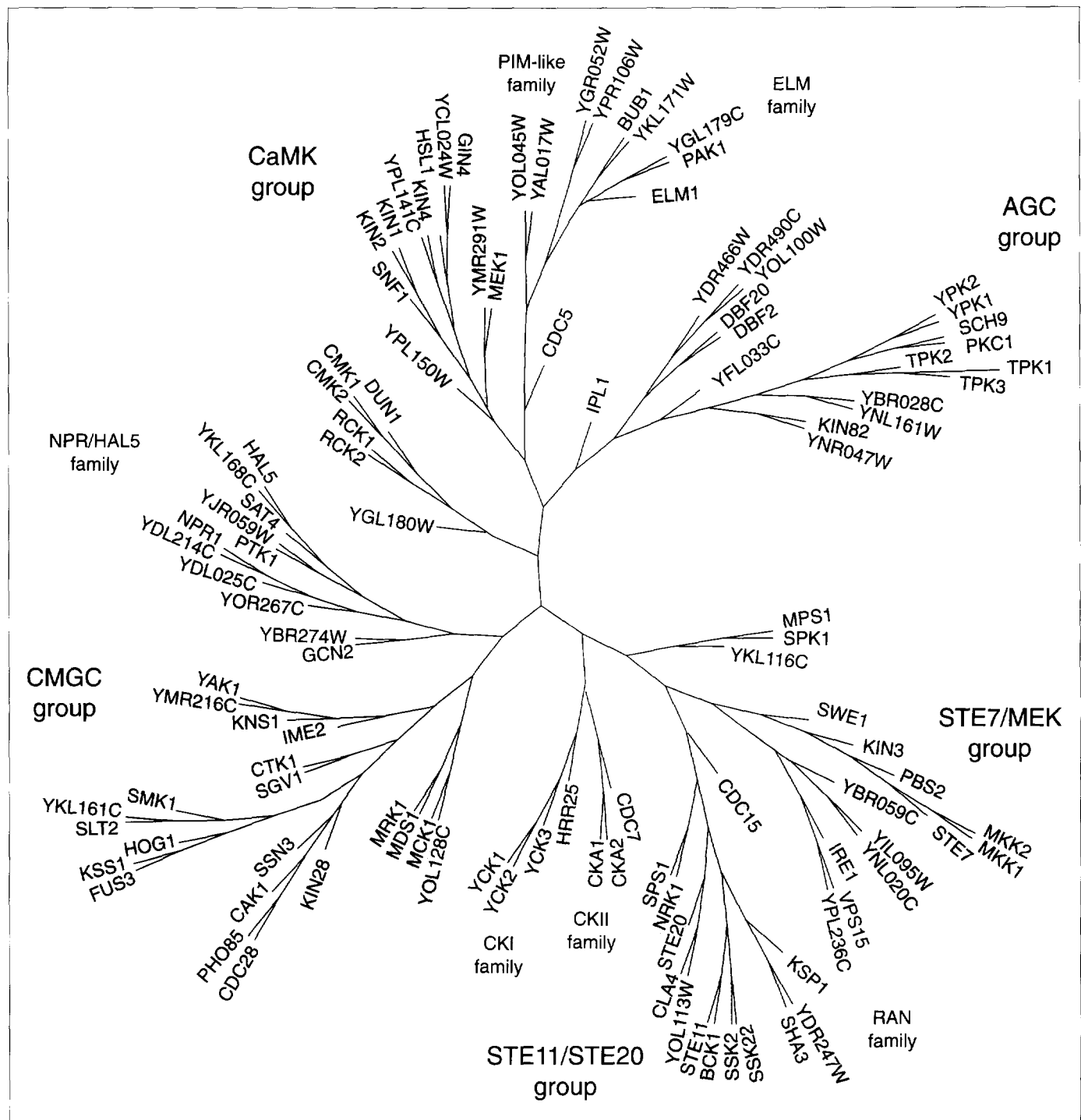
Figure 1

Dendrogram of the budding yeast protein kinase superfamily. The catalytic domains of 113 yeast protein kinases were aligned using SAM[39], a multiple sequence alignment program that applies a linear hidden Markov model to facilitate recognition of conserved subdomains within a protein family (http://www.cse.ucsc.edu/research/compbio/sam.html). The SAM alignment was run on a MasPar parallel computer and the results were inputted into PROTPARS, a protein sequence parsimony method, to build an unrooted phylogeny. PROTPARS is part of the PHYLIP package written by J. Felsenstein of the University of Washington (http://evolution.genetics.washington.edu/phylip.html).

that Ste20 and Cla4 share a function in this process. Nevertheless, the double *ste20/cla4* mutant can still assemble actin, a process regulated by Cdc42 and other Rho family members. This could be accounted for by YOL113W, a third member of the Ste20 family whose existence has been revealed by the genome sequencing project; it will obviously be important to analyse the function of YOL113W and determine to what extent its function is redundant with those of Ste20 and Cla4. Like the other two family members, YOL113W has both a phospholipid-binding pleckstrin homology (PH) domain and a Rac/Cdc42-binding motif $(SxPx_{4-6}HxxH)$ upstream of its catalytic domain.

The function of Ste20 subfamily protein kinases is of significant interest, because it expanded greatly during evolution and there are a large number of Ste20-related protein kinases in vertebrates, such as the Paks, which are regulated by Cdc42 and other members of the Cdc42 GTPase family.

Cell-cycle control

The regulation of the cell cycle involves many types of protein kinase. In

19

budding yeast, Cdc28 is the only known CDK with an essential role in cell-cycle regulation, although two other protein kinases, Pho85 and Kin28, bind and are activated by cyclins and might have roles in cell-cycle progression[13]. Pho85 can play a direct role in G1 regulation, whereas Kin28 has an indirect role as a component of the basal transcription factor TFIIH. All three of these yeast protein kinases contain a canonical AIRE sequence in catalytic subdomain III, which is part of the α1-helix that interacts through the conserved Ile with the cyclin subunit.

An additional member of the CDK cluster, Ssn3/Srb10, has no apparent role in cell-cycle regulation, but forms a complex with cyclin Srb11 as part of the larger RNA polymerase II holoenzyme[13]. Ctk1 is a divergent CDK identified as an RNA polymerase II carboxy-terminal domain kinase, which binds Ctk2, a cyclin-related protein[13], and contains SIRE instead of AIRE. The genome sequence revealed a number of new cyclins, but only one new CDK-related gene, CAK1/CIV1/YFL029C. This gene is unlikely to encode a true CDK, because it has an AKFE sequence instead of the AIRE motif; indeed CAK1/CIV1/YFL029C has recently been shown to be active as a monomer and to be a CDK-activating kinase (CAK) that phosphorylates Thr in the activation loop of Cdc28 (Refs 14–16), a function that in mammalian cells is currently thought to be carried out by cyclin H/Cdk7. The CDK subfamily evidently underwent significant expansion during the evolution of the multicellular eukaroytes, as in vertebrates at least four CDKs are directly involved in cell-cycle regulation and eight CDKs are known altogether.

Another cell-cycle regulatory protein kinase, NIMA, which is required for the G2–M transition in Aspergillus, has so far been only identified in filamentous fungi[17]. Given the highly conserved nature of cell-cycle regulation, one might anticipate that other eukaryotes would have NIMA homologues, and some evidence for a NIMA response pathway in vertebrates has been obtained[18,19]. However, no true homologues have been identified in other eukaryotes, although there are several NIMA-related protein kinases in vertebrates. The yeast Kin3/Npk1 protein kinase is quite closely related to NIMA in its catalytic domain[20,21], but lacks the carboxy-terminal regulatory domain, which is critical for NIMA cell-cycle function. Thus, it appears that there might be no budding yeast homologue of NIMA.

There are several solitary yeast cell-cycle regulatory protein kinases, including Cdc5, Cdc7 and Cdc15. Cdc5 is required for exit from mitosis[22], and its three counterparts in higher vertebrates, Plk, Fnk/Prk and Sak, form a subfamily. Plk, Polo (a Drosophila relative) and Plo1 (a fission yeast relative), like Cdc5, have been implicated in progression through mitosis. However, Cdc7, a protein kinase in the casein kinase II family, which is required for initiation of DNA synthesis during S phase[23] and is related to Hsk1 from fission yeast, has no known mammalian counterpart. Cdc15, which is essential for completion of mitosis[24], also has no known mammalian homologue; it is related to human MESS1 in the cata-lytic domain, but has a long dissimilar carboxy-terminal tail.

Diversity of yeast protein kinases

Most of the main vertebrate subfamilies of protein kinases are represented in yeast. For example, in the AGC group there are multiple cAMP-dependent protein kinases (PKAs), a single protein kinase C (PKC), and 70 kDa S6 kinase-related protein kinases, but no protein kinases closely related to cGMP-dependent protein kinase, Rsk or βARK. In the CaMK group, there are Ca^{2+}–calmodulin-regulated and AMP-dependent protein kinases, but no true MLCK, perhaps because myosin-based motility is limited in yeast. In the CMGC group, all the main subfamilies are represented in yeast including CDKs, MAPKs, and GSK3- and Clk-related protein kinases.

New groups. As this is the first analysis of all the protein kinases present in a complete eukaryotic genome, it is appropriate to designate two new groups that are conserved in all eukaryotes, namely the STE11/STE20 group including the MAPKKKs and the STE7 group including the MAPKKs. In addition, several new subfamilies can be established in the 'Other group', including casein kinase I and II, the Ran subfamily and two subfamilies that appear to be unique to budding yeast, Npr/Hal5 and Elm subfamilies. Notably absent, however, are receptor-type protein kinases, and Raf-related protein kinases, which are MAPKKKs that act downstream of receptor protein-tyrosine kinases. Although yeast has several STE11/ MEKK-family MAPKKKs, the absence of Raf-related MAPKKKs and receptor protein kinases in general might reflect the fact that yeast has little need for intercellular communication, other than to respond to mating pheromones, which is

accomplished by G protein-coupled receptors. Many of the protein kinase subfamilies have undergone significant expansion during evolution; for example, there is only one PKC in yeast, but at least nine in vertebrates.

No true protein-tyrosine kinases. As anticipated from many unfruitful sequence-based searches, budding yeast has no members of the true protein-tyrosine kinase family. The progenitor for this family probably arose when multicellular organisms evolved. The driving force behind the evolution of protein-tyrosine kinases was presumably the need for a signaling mechanism for cell–cell communication within a multicellular organism. The concomitant evolution of phosphotyrosine-binding domains that could mediate protein-tyrosine kinase signal-dependent protein–protein interactions must also have been a critical event.

The absence of typical protein-tyrosine kinases, however, does not mean that enzymes of this specificity are completely lacking, and there are several examples of what have been termed dedicated protein-tyrosine kinases. For example, Swe1, a member of the Wee1 family, phosphorylates Tyr19 in Cdc28, negatively regulating the activity of this CDK[25]. Other Wee1 family members can autophosphorylate on serine, threonine and tyrosine, and as a result are commonly known as dual-specificity protein kinases. Although there are no other Swe1-related protein kinases in yeast that could regulate CDK function, there are two protein kinases, Spk1/Rad53/Mec2/Sad1 and Mps1/Rpk1, involved in S and G2 checkpoint control and spindle pole body duplication respectively. These kinases are also dual-specificity protein kinases that, at least in vitro, have protein-tyrosine kinase activity[26,27]. Yeast also has true dual-specificity protein kinases in the MAPKK family, which phosphorylate the Thr and Tyr in a TxY motif in the activation loop of members of the MAP kinase family. The existence of three bona fide protein-tyrosine phosphatases in yeast underscores the importance of protein-tyrosine phosphorylation in this organism.

All the protein kinases classified in the protein-tyrosine kinase family based on sequence analysis have experimentally verified tyrosine phosphorylating specificity. However, it is not yet known exactly how tyrosine, rather than serine or threonine, is selected for phosphorylation, despite the availability of several protein-serine and protein-tyrosine kinase

catalytic-domain three-dimensional structures. This means that one cannot exclude the possibility that a novel solitary protein kinase is not a protein-tyrosine kinase until it is tested biochemically. Indeed, mammals have a novel protein-tyrosine kinase, A6, which is totally unrelated in sequence to the conventional protein-tyrosine kinase family[28]. There is an A6-related gene, YGR080W, in yeast, and additional homologues in the human EST database. It will be interesting to test whether these genes encode protein-tyrosine kinases.

Protein kinases found only in yeast

Some subfamilies of yeast protein kinases so far appear to be unique to budding yeast. These include a PKA-related family (YOL100W, YDR490C, YDR466W), a Nek-like family (YNL020, YIL095W, YBR059C) and the Npr/Hal5 family, which has nine members divided into two groups related to Hal5, a protein kinase involved in salt and pH tolerance, and Npr1, a protein kinase involved in regulating transport systems for nitrogen nutrients under conditions of catabolite derepression[29]. If one were to predict what sort of protein kinases might be unique to yeast, those involved in nutrient uptake or resistance to environmental stress would be obvious candidates. The same might be true for the Elm family, where the eponymous protein kinase Elm1 is involved in pseudohyphal growth[30], a process that is unique to yeast. There are also a number of yeast protein kinases that have homologues in other species, but not in vertebrates; for instance the Ran family contains protein kinases closest to the fission yeast Ran1 protein kinase.

There are currently four yeast protein kinases that have no known homologues in other species (Bub1, YKL171W, YGR052W and YPR106W). However, although protein kinases in these families have not been identified in vertebrates so far, several of the yeast protein kinases are close relatives of *Caenorhabditis elegans* protein kinases, where the genome sequence is now nearly 60% complete. Thus, as the human genome sequence progresses, one can anticipate that many of the apparently unique yeast protein kinases will prove to have vertebrate homologues.

Non-conventional protein kinases

In addition to genes in the conventional protein kinase superfamily, there is also a single gene, that encodes a protein in the prokaryotic 'histidine'

protein kinase family, Sln1 (Ref. 31). Such proteins autophosphorylate on a histidine residue in response to a specific stimulus, and then this phosphate is transferred to an acceptor signal response protein on an aspartate residue. Given the frequency of such protein kinases in prokaryotes, where more than ten are known in *Escherichia coli*, it is surprising that there are not more protein kinases of this type in yeast. It is interesting, however, that the Sln1 histidine kinase plays a role in the response of yeast cells to osmotic pressure, a response that it is transduced in bacterial cells by the histidine kinase, EnvZ[31]. The YIL042C protein is related to the mitochondrial branched chain α-ketoacid and pyruvate dehydrogenase protein kinases, which are unusual protein-serine kinases that are recognizable members of the 'histidine' protein kinase family[32,33].

There are a number of eukaryotic proteins with bona fide protein kinase activity, such as the *Dictyostelium* myosin heavy chain protein kinase[34], that are structurally unrelated to either the eukaryotic protein kinase superfamily or the prokaryotic signal response protein kinases. Well over 60% of the genes in the yeast genome are of unknown function, and it is certainly possible that some of them will be unconventional protein kinases, such as the A6-related protein discussed above.

The yeast genome has ten genes in the lipid kinase family, which have a domain related to the catalytic domain of the protein kinase superfamily. Some of these are bona fide phosphatidylinositol kinases (e.g. Vps34, Pik1, etc.), whereas others have not been found to have lipid kinase activity. Given that DNA PK, a mammalian protein in this family, is a genuine protein kinase whose activity is stimulated by double-stranded DNA ends, there has been speculation that some of the members of the lipid kinase family are in fact protein kinases[35]. Indeed, it appears likely that the members of this family that are involved in checkpoint function, such as Mec1 and Tel1, are protein kinases. Moreover, another lipid kinase subfamily, which includes rapamycin-binding proteins, Tor1 and Tor2, is known to autophosphorylate, and it is likely that these are also protein kinases, although their salient substrates remain to be identified.

Perspectives

If we include all the different types of protein kinase encoded by the yeast genome, we reach a total of ~120; a

number that is a little lower than the most recent estimate of the number of yeast protein kinases, which was based on the sequencing of chromosome III (Ref. 36). Recent analysis of GenBank and unfinished sequence databases for *C. elegans* (kindly provided by the Sanger Centre, Cambridge, UK) so far reveals 270 unique protein kinases (G. D. Plowman, unpublished). However, although these data represent approximately 60% of the 100 Mb nematode genome, which is about eight times larger than that of yeast, the number of kinases per kb of DNA is only half that predicted from the analysis of the yeast genome. However, the number and length of introns in higher eukaryotes is much greater than in yeast, which decreases the percentage of coding regions.

By extrapolation, we can estimate the number of protein kinase genes in mammals, which have about four times as many genes as *C. elegans*. Based on this estimation, a prediction of more than 1000 protein kinase genes in the human genome still seems a reasonable one[37], particularly if one takes into account the expansion of the protein-tyrosine kinases used for intercellular signalling in higher organisms. By the time the human genome project is completed we will know how accurate this estimate was.

References

1 Garrels, J. I. (1995) *Nucleic Acids Res.* 24, 46–49
2 Hieter, P., Bassett, D. E., Jr and Valle, D. (1996) *Nat. Genet.* 13, 253–255
3 Johnston, M. (1996) *Curr. Biol.* 6, 500–503
4 Waskiewicz, A. J. and Cooper, J. A. (1995) *Curr. Biol.* 7, 798–805
5 Liu, H., Styles, C. A. and Fink, G. R. (1993) *Science* 262, 1741–1744
6 Sia, R. A. and Mitchell, A. P. (1995) *Mol. Cell. Biol.* 15, 5279–5287
7 Krisak, L. *et al.* (1994) *Genes Dev.* 8, 2151–2161
8 Friesen, H., Lunz, R., Doyle, S. and Segall, J. (1994) *Genes Dev.* 8, 2162–2175
9 Simon, M. N. *et al.* (1995) *Nature* 376, 702–705
10 Ramer, S. W. and Davis, R. W. (1993) *Proc. Natl. Acad. Sci. U. S. A.* 90, 452–456
11 Leberer, E. *et al.* (1992) *EMBO J.* 11, 4815–4824
12 Cvrckova, F. *et al.* (1995) *Genes Dev.* 9, 1817–1830
13 Poon, R. Y. C. and Hunter, T. (1995) *Curr. Biol.* 5, 1243–1247
14 Espinoza, F. H. *et al.* (1996) *Science* 273, 1714–1717

15 Kaldis, P., Sutton, A. and Solomon, M. J. (1996)
   Cell 86, 553–564
16 Thuret, J-Y., Valay, J-G., Faye, G. and Mann, C.
   (1996) *Cell* 86, 565–576
17 Osmani, S. A. and Ye, X. X. (1996) *Biochem. J.*
   317, 633–641
18 Lu, K. P. and Hunter, T. (1995) *Cell* 81,413–424
19 O'Connell, M. J., Norbury, C. and Nurse, P.
   (1994) *EMBO J.* 13, 4926–4937
20 Kambouris, N. G., Burke, D. J. and Creutz, C. E.
   (1993) *Yeast* 9, 141–150
21 Schweitzer, B. and Philippsen, P. (1992) *Mol.
   Gen. Genet.* 234, 164–167
22 Kitada, K., Johnson, A. L., Johnston, L. H. and
   Sugino, A. (1993) *Mol. Cell. Biol.* 13,
   4445–4457

23 Patterson, M., Sclafani, R. A., Fangman, W. L.
   and Rosamond, J. (1986) *Mol. Cell. Biol.* 6,
   1590–1598
24 Schweitzer, B. and Philippsen, P. (1991) *Yeast*
   7, 265–273
25 Booher, R. N., Deshaies, R. J. and
   Kirschner, M. W. (1993) *EMBO J.* 12,
   3417–3426
26 Lauze, E. *et al.* (1995) *EMBO J.* 14, 1655–1663
27 Zheng, P. *et al.* (1993) *Mol. Cell. Biol.* 13,
   5829–5842
28 Beeler, J. F. *et al.* (1994) *Mol. Cell. Biol.* 14,
   982–988
29 Vandenbol, M., Jauniaux, J. C. and Grenson, M.
   (1990) *Mol. Gen. Genet.* 222, 393–399
30 Blacketer, M. J. *et al.* (1993) *Mol. Cell. Biol.*

   13, 5567–5581
31 Posas, F. *et al.* (1996) *Cell* 86, 865–875
32 Popov, K. M. *et al.* (1992) *J. Biol. Chem.* 267,
   13127–13130
33 Popov, K. M. *et al.* (1993) *J. Biol. Chem.* 268,
   26602–26606
34 Futey, L. M., Medley, Q. G., Cote, G. P. and
   Egelhoff, T. T. (1995) *J. Biol. Chem.* 270,
   523–529
35 Hunter, T. (1995) *Cell* 83, 1–4
36 Hunter, T. (1994) *Semin. Cell Biol.* 5, 367–376
37 Hunter, T. (1987) *Cell* 50, 823–829
38 Hanks, S. K. and Hunter, T. (1995) *FASEB J.*
   9, 576–596
39 Krogh, A. *et al.* (1994) *J. Mol. Biol.* 235,
   1501–1531

Protein architecture, dynamics and allostery in tryptophan synthase channeling

Peng Pan, Eilika Woehl and Michael F. Dunn

The $\alpha_2\beta_2$ form of the tryptophan synthase bienzyme complex catalyses the last two steps in the synthesis of L-tryptophan, consecutive processes that depend on the channeling of the common metabolite, indole, between the sites of the α- and β-subunits through a 25 Å-long tunnel. The channeling of indole and the coupling of the activities of the two sites are controlled by allosteric signals derived from covalent transformations at the β-site that switch the enzyme between an open, low-activity state, to which ligands bind, and a closed, high-activity state, which prevents the escape of indole.

THE PHENOMENON OF direct metabolite transfer between sequential enzyme pairs in a metabolic cycle is classified as substrate channeling[1]. The tryptophan synthases from enteric bacteria, with subunit composition $\alpha_2\beta_2$, are the best-characterized examples of substrate-channeling, multienzyme complexes[2,3]. These enzymes catalyse the last two steps (Fig. 1) in the biosynthesis of L-tryptophan (L-Trp). The α-subunit has an $(\alpha/\beta)_8$-barrel folded motif, and catalyses the cleavage of 3-indole-D-glycerol 3'-phosphate (IGP) to indole and D-glyceraldehyde-3'-phosphate (G3P). The β-subunit is a pyridoxal phosphate-requiring enzyme that catalyses the conversion of L-serine (L-Ser) and indole to L-Trp and a water molecule[2]. The

β-reaction occurs in two stages; in Stage I, L-Ser reacts with enzyme-bound pyridoxal 5'-phosphate (PLP) to form the quasi-stable α-aminoacrylate intermediate, E(A-A), the species poised for reaction with indole; in Stage II, indole reacts with E(A-A) to form L-Trp (Fig. 1). Efficiency is achieved by channeling the product of the first enzyme (indole) to the second enzyme from the α-site to the β-site through a 25 Å-long tunnel (Fig. 2).

Here, we review recent findings showing that, in the overall catalytic cycle, β-site covalent reactions with substrate trigger allosteric signals that flip the enzyme between open (low activity) and closed (high activity) conformations. These events serve two functions: (1) conversion to the closed state prevents the escape of indole, and (2) switching between activity states couples the catalytic cycles of the two enzymes.

P. Pan, E. Woehl and **M. F. Dunn** are at the Department of Biochemistry, University of California, Riverside, CA 92521-0129, USA.

Physical and dynamic constraints are dictated by function

The elegant crystallographic work of Hyde *et al.*[4] and Rhee *et al.*[5] on the *Salmonella typhimurium* tryptophan synthase bienzyme complex $(\alpha_2\beta_2)$ has contributed significant advances to our knowledge of structure–function relationships in substrate channeling by multienzyme complexes. Their efforts have provided the first example showing the three-dimensional structure of the molecular machinery required for channeling in a stable multienzyme complex. Out of this structural work and from recent mechanistic studies[6–12], there has emerged the realization that, for efficient substrate channeling to occur between enzyme pairs, a rather stringent set of physical and dynamic constraints must be met. Those evident in the tryptophan synthase example so far are as follows: (1) The architecture of the multienzyme complex must provide a physical structure with dynamic properties that constrain the degrees of freedom of the common metabolite so that transfer from one site to the next is assured. (2) Catalysis at the two active sites must be coupled such that turnover at each site occurs in phase.

The structure[4] of $\alpha_2\beta_2$ partially explains how the first constraint is satisfied; the α- and β-sites of each heterologous dimer are connected by a 25 Å-long tunnel (Fig. 2). The kinetic studies of Dunn *et al.*[6], Lane and Kirschner[7] and Anderson *et al.*[8] have established that the tunnel actually functions as the conduit for the transfer of indole between sites.

An interconnecting tunnel is insufficient to ensure channeling

Two additional criteria must be met to achieve efficient phasing of the vectoral transfer of indole with the β-site chemistry (Fig. 1) so that L-Trp is efficiently