

© Springer-Verlag New York Inc. 1997

Experimental Indication in Favor of the Introns-Late Theory: The Receptor Tyrosine Kinase Gene from the Sponge *Geodia cydonium*

Vera Gamulin,¹ Alexander Skorokhod,^{2,3} Vadim Kavsan,² Isabel M. Müller,³ Werner E. G. Müller³

¹ Department for Molecular Genetics, Institute Ruder Boskovic, 10000 Zagreb, Croatia

² Department of Biosynthesis of Nucleic Acids, Institute of Molecular Biology and Genetics, Ukrainian Academy of Sciences,

252627 Kiev, Ukraine

³ Institut für Physiologische Chemie, Abteilung Angewandte Molekularbiologie, Universität, Duesbergweg 6, D-55099 Mainz, Germany

Received: 8 August 1996 / Accepted: 4 November 1996

Abstract. We have analyzed the gene that encodes receptor tyrosine kinase (RTK) from the marine sponge Geodia cydonium, which belongs to the most ancient and simple metazoan groups, the Porifera. RTKs are enzymes found only in metazoa. The sponge gene contains two introns in the extracellular part of the protein. However, the rest of the protein (transmembrane and intracellular part), including the tyrosine kinase (TK)domain, is encoded by a single exon. In contrast, all TK genes, so far known only from higher animals (vertebrates), contain several introns especially in the TKdomain. The TK-domain of G. cydonium shows similarity with numerous members of receptor as well as nonreceptor TKs. Phylogenetic analysis of the sponge TK-domain indicates that this enzyme branched off first from the common tree of metazoan TK proteins. Consequently, we assume that introns, found in the TKdomains of genes from higher animals, were inserted into these genes after splitting off the sponge taxa from other metazoan organisms (over 600 million years ago). Our results support the view that ancient genes were not "in pieces."

Key words: *Geodia cydonium* — Sponge — Metazoan protein molecules

Introduction

Protein tyrosine kinases (PTKs) are a large group of enzymes that specifically phosphorylate tyrosine residues (Hardie and Hanks 1995). They play important roles in the response of cells to different extracellular stimuli and are essential proteins most notable for control of growth and differentiation (Hunter et al. 1992). Many PTKs serve as receptors and signal transducers for circulating peptide hormones and growth factors (Stoddard et al. 1992). PTKs were first discovered in oncogenic retroviruses and subsequently identified and analyzed in a variety of different metazoan organisms (Hanks and Quinn 1991). Hundreds of PTK primary structures are known; all have been isolated from metazoan organisms (Hardie and Hanks 1995). PTKs together with serine/ threonine kinases constitute the largest known protein superfamily (Hardie and Hanks 1995).

All PTKs possess a closely related tyrosine kinase (TK)-domain which is specific for the phosphorylation of tyrosine only (Hunter et al. 1992; Ullrich and Schlessinger 1990). It was estimated by Hunter (1987) that the number of protein kinases, including TKs, present in a highly evolved metazoan organism might be 1,000 or

The sequences reported in this paper have been deposited in the EMBL database (cDNA for receptor tyrosine kinsase from *G. cydonium*, accession No. X77528; for the gene, No. X94128) *Correspondence to:* W.E.G. Müller

GCTKGe	gttgctggtatcacagctgcacactggatgaaaaacactcacgtcaatatcataatgtagtatgtat	90
GCTKGe	TGGACT <u>CCTATATGG</u> CTGTATTACTCTGTACAGAAGTGTGTTTGTAGTGTCTCTAATGCATAGCTAATAATGGTGTAGGCGTCAGATGAT	180
GCTKGe	GTGTCTTTCGAGCATCTGTTGTTATAGTAGCTATGGTGGCTATGGTGCAGAAATGGGTCTAAAAAAGGTTTGTTT	270
GCTKGe	tctgatgctctgcctagctgggttgcttgttttatcccatgtttcgtcatacaaagcacccactcttttactattccccacgcctcaattt	360
GCTKGe	ATAGTCTOTGTACTTOTCACCAGGAACGCAACCTCTTACCTGTAATATTTTATACCAGTAGTGTGTACATGTGACCATCTGTAGTATACTGT	450
GCTRGE	ATATATATATATCTUGGCATTGACTTGACTTTAATATTATTATGGCATAGGAATGGTAGGACAATGGTAGGACAATGGTAGGAGGAGAGAGA	540
GUINGE		0.50
GCTKGe	TTGCAGGTGTGTTCC <u>TATTAA</u> GACTCCACAGTAGATACAGCAGCATGGTTGATTGAGTACATATTGTATGTA	720
GCTKGe	GAACAAATTTTCTAAGAACAAAATATCTTTCCACCTTGCCATTGTGTTTTGCAGAGGAGGAAGTATATACACTATCCCTGATAGAGAGGAGGAAA	810
GCTKGe	CGGGAAATCTTCCACTAATTTTACAGTTGTGTGTGTGTGCATTTTGAAGTTGGATTGTGTTTATTACTTTTGTTTTGCCCCTTCATGTAAGTAA	900
GCTKGe	ATTTCTTTCTTTGTTCAACAGATTTGGACACAGGCTCTGTCTTCATTGTCTTCCTTC	990
GCTKGe	CCAGAGAGGGTAAAGAAATTGGGCTATCGTGGGTTTCTTTTAGAGCATGGTTTTCTTCTTCTCAAGGTTCTTTCAGGGCTTCAAGATGTT	1080
GCTKGe	GGACCTACAGTACCTCCAGTGACATACAATGGAATTGTGCCCAAGAATATCACCATGTGGAGCCTCTCCAACAAGTTTGAATTGCAAGGCA	1170
GCTKGe	GGAGAAGGAGTGTTGAGTGATGGAGAACATTGGACCTACAGACTCAGTGGATGTCAGTGATCCAGGACCAGGTCACCAGGTTCTTTGTCTGG	1260
GCTKGe	METAlaLeuTrpHisTrpCysSerLeuArgLeuMetSerProValSerProThrLeuMetHisTyrPheLeuSerLeuPr CATAGAGACA <u>ATG</u> GCTCTGTGGCACTGGTGTTCACTAAGATTGATGTCTCCCAGTGTCACCTACATTGATGCAT <u>T</u> ACTTCTTGAGTCTTCC	27 1350
	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	57
GCTKGe	OSEFAIASEFITERGLEUPFOGLUTHTHEVAITTETTETHEFNERSDITETHFFFOFFORGVALGIUTHTHINVAITHFFOGLUSECY TTCTGCTAGTATACGACTACCTGAGACTACTGTTATTATCACTGATATCACTCCACCAAGAGTTGAGACTACTGTCACCCCTGAGAGCTG	1440
	sSerPheSerSerThrSerAsnThrLeuThrArgAsnThrTyrThrValSerGlnAsnProAspAspIleLeuIleThrPheAsnPheAs	87
GCTKGe	TTCTTTCTCCCCCCACTAGTAATACACTGACTAGGAACACGTATACAGTATCACAGAATCCAGATGATATTCTTATCACATTCAATTCAA	1530
00000-	nGinSerThrAspTrpMetPhelleSerGluileLeuCeySAlaGlyAspProProSerSerIleSerCySAspSerProThrThrAs	1620
GCIKGe		1020
	nProThrThrGlnThrProThrThrSerProGlvProSerProThrProProSerLeuThrLeuSerSerProProProThrGlvLeuPr	147
GCTKGe	CCCCACAGACACCACCACCCCCCCCCCCCCCCCCCCCC	1710
GCTKGe	oValSerProAspLeuSerGlnProHisSerValThrLeuThrCysSerAlaAlaSerProProAlaThrGlyTyrGlnTyrGlnTrpGl AGTGAGCCCTGACCTGAGCCAGCCACACTCTGTCACTCTCACTTGCTCTGCCGCCAGTCCTCCTGCCACTGGCTACCAATACCAGTGGCA	177 1800
	${\tt nTrpArgArgHisLvsThrLeuLeuSerAsnThrThrArgPheSerIleThrProSerThrAsnThrGlnSerSerLeuVallleSe}$	207
GCTKGe	GTGGAGGAGGCATGGGACACTACTGAGCAACACCACTAGATTCTCTATCACACCCCCCAACACACTCAGTCCAGTAGTCTAGTCATATC	1890
GCTKCD	AGATTCTCTATCACACCCCCCCCACCAACACTCAGTCCAGTAGTCTAGTCATATC	
	${\tt ArgPheSerIleThrProSerThrAsnThrGlnSerSerLeuValIleSe}$	
	ArgPheSerIleThrProSerThrAsnThrGlnSerSerSerLeuValIleSe	0.9.7
	ArgPheSerIleThrProSerThrAsnThrGlnSerSerSerLeuVallleSe rGlyLeuArgTyr <u>Ser</u> AspAlaGlyAspTyrMetCysThrValLysTyrGly <u>Ala</u> CysPro <u>Gly</u> GlyValAspCysSerGlyThrThrPr	237
GCTKGe	ArgPheSerIleThrProSerThrAsnThrGlnSerSerSerLeuVallleSe rGlyLeuArgTyr <u>Ser</u> AspAlaGlyAspTyrMetCysThrValLysTyrGly <u>Ala</u> CysPro <u>Gly</u> GlyValAspCysSerGlyThrThrPr TGGTCTCAGATATTCTGATGCAGGAGACTACATGTGTACAGTGGAGTATGGAGCATGTCCTGGTGGAGGGGGGGG	237 1980
GCTKGe	ArgPheSerIleThrProSerThrAsnThrGlnSerSerSerLeuVallleSe      rGlyLeuArgTyr <u>Ser</u> AspAlaGlyAspTyrMetCysThrValLysTyrGly <u>Ala</u> CysPro <u>Gly</u> GlyValAspCysSerGlyThrThrPr      TGGTCTCAGATATTCTGATGCAGGAGACTACATGTGTACAGTGGAGGAGGAGGAGCAGGACTGCCTGGTGGAGGGGGGGG	237 1980
GCTKGe GCTKcD	ArgPheSerIleThrProSerThrAsnThrGlnSerSerSerLeuVallleSe      rGlyLeuArgTyrSerAspAlaGlyAspTyrMetCysThrValLysTyrGlyAlaCysProGlyGlyValAspCysSerGlyThrThrPr      TGGTCTCAGATATTCTGATGCAGGAGACTACATGTGTACAGTGGAGTGGAGCAGTGGAGGAGGAGGACGACGACCACTCC	237 1980
GCTKGe GCTKcD	ArgPheSerIleThrProSerThrAsnThrGlnSerSerSerLeuVallleSe    rGlyLeuArgTyrSerAspAlaGlyAspTyrMetCysThrValLysTyrGlyAlaCysProGlyGlyValAspCysSerGlyThrThrPr    TGGTCTCAGATATTCTGATGCAGGAGACTACATGTGTACAGTGGAGTGGAGGAGCATGTCCTGGTGGAGTGGAGGGGGACGACCACACCC	237 1980
GCTKGe GCTKcD	$\label{eq:light} ArgPheSerIleThrProSerThrAsnThrGlnSerSerSerLeuVallleSerGlyLeuArgTyrSerAspAlaGlyAspTyrMetCysThrValLysTyrGlyAlaCysProGlyGlyValAspCysSerGlyThrThrPrTGGTCTCAGATATTCTGATGCAGGAGACTACATGTGTACAGTGGAGTAGGAGCATGTCCTGGTGGAGTGGAGGGGGGACGACAACTCCCTGGTGGAGTGGAGGGGGGGG$	237 1980 247
GCTKGe GCTKCD GCTKGe	$\label{eq:light} ArgPheSerIleThrProSerThrAsnThrGlnSerSerSerLeuVallleSerGlyLeuArgTyrSerAspAlaGlyAspTyrMetCysThrValLysTyrGlyAlaCysProGlyGlyValAspCysSerGlyThrThrPrTGGTCTCAGATATTCTGATGCAGGAGACTACATGTGTACAGTGGAGGAGGACTACATGTGCAGGGGGGGG$	237 1980 247 2070
GCTKGe GCTKCD GCTKGe	$\label{eq:light} ArgPheSerIleThrProSerThrAsnThrGlnSerSerSerLeuVallleSerGlyLeuArgTyrSerAspAlaGlyAspTyrMetCysThrValLysTyrGlyAlaCysProGlyGlyValAspCysSerGlyThrThrPrTGGTCTCAGATATTCTGATGCAGGAGACTACATGTGTACAGTGGAGTAGGAGGACTGCCTGGTGGAGGGGGGGG$	237 1980 247 2070
GCTKGe GCTKCD GCTKGe GCTKCD	$\label{eq:light} ArgPheSerIleThrProSerThrAsnThrGlnSerSerSerLeuValIleSerGlyLeuArgTyrSerAspAlaGlyAspTyrMetCysThrValLysTyrGlyAlaCysProGlyGlyValAspCysSerGlyThrThrPrTGGTCTCAGATATTCTGATGCAGGAGACTACATGTGTACAGTGGAGTAGGAGGACTACCTGGAGGAGGACTGCCTGGTGGAGGAGGACTGCAGGGAGGACTACATGTGTACAGTGGAGGAGCATGCCTGGTGGAGGGGGGGG$	237 1980 247 2070
GCTKGe GCTKCD GCTKGe GCTKCD	ArgPheSerIleThrProSerThrAsnThrGlnSerSerSerLeuValIleSe    rGlyLeuArgTyrSerAspAlaGlyAspTyrMetCysThrValLysTyrGlyAlaCysProGLyGlyValAspCysSerGlyThrThrPr    TGGGTCTCAGATATTCTGATGCAGGAGACTACATGTGTACAGTGGAGTAGGAGTATGGAGCATGTCCTGGTGGAGTGGACTGCAGTGGAACAACTCC    IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	237 1980 247 2070
GCTKGe GCTKGe GCTKCD GCTKGe	$\label{eq:relation} \begin{tabular}{lllllllllllllllllllllllllllllllllll$	237 1980 247 2070 2160
GCTKGe GCTKCD GCTKCC GCTKCD GCTKGe	$\label{eq:starter} ArgPheSerIleThrProSerThrAsnThrGlnSerSerSerLeuValIleSerGlyLeuArgTyrSerAspAlaGlyAspTyrMetCysThrValLysTyrGlyAlaCysProGlyGlyValAspCysSerGlyThrThrPrTgGTCTCAGATATTCTGATGCAGGAGACTACATGTGTACAGTGGAGTATGGAGCATGTCCTGGTGGAGTGGACTGCAGTGGAACAACTCCGIUIUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU$	237 1980 247 2070 2160
GCTKGe GCTKCD GCTKCC GCTKCD GCTKGe	$\label{eq:relation} \begin{tabular}{lllllllllllllllllllllllllllllllllll$	237 1980 247 2070 2160 273
GCTKGe GCTKCD GCTKCD GCTKGe GCTKGe	$\label{eq:light} \\ \begin{tabular}{lllllllllllllllllllllllllllllllllll$	237 1980 247 2070 2160 273 2250
GCTKGe GCTKCD GCTKCD GCTKGe GCTKGe	ArgPheSerIleThrProSerThrAsnThrGlnSerSerSerLeuValIleSe    rGlyLeuArgTyrSerAspAlaGlyAspTyrMetCysThrValLysTyrGlyAlaCysProGlyGlyValAspCysSerGlyThrThrPr    TGGTCTCAGATATTCTGATGCAGGAGACTACATGTGTACAGTGGAGGAGATGGGAGCATGTCCTGGTGGAGTGGACTGCAGTGGAACAACTCC	237 1980 247 2070 2160 273 2250
GCTKGe GCTKCD GCTKCD GCTKGe GCTKGe GCTKCD	ArgPheSerIleThrProSerThrAsnThrGlnSerSerSerLeuValIleSe    rGlyLeuArgTyrSerAspAlaGlyAspTyrMetCysThrValLysTyrGlyAlaCysProGlyGlyValAspCysSerGlyThrThrPr    TGGTCTCAGATATTCTGATGCAGGAGACTACATGTGTACAGTGGAGGAGGACTAGCAGGAGGACTGCCTGGTGGAGGGGGGGG	237 1980 247 2070 2160 273 2250
GCTKGe GCTKCD GCTKCD GCTKGe GCTKGe GCTKCD	$\label{eq:start} ArgPheSerIleThrProSerThrAsnThrGlnSerSerSerLeuVallleSerGlyThrThrPrTGGTCCAGATGTTCTGATGCAGGAGAGATCCTGTTCTGATGTACAGTGGAGGAGTGTGGAGGAGGAGAGAGGGGGGGG$	237 1980 247 2070 2160 273 2250
GCTKGe GCTKCD GCTKCD GCTKGe GCTKGe GCTKCD	ArgPheSerIleThrProSerThrAsnThrGlnSerSerSerLeuVallleSe    rGlyLeuArgTyrSerAspAlaGlyAspTyrMetCysThrValLysTyrGlyAlaCysProGlyGlyValAspCysSerGlyThrThrPr    TGGTCTCAGATATTCTGATGCAGGAGACTACATGTGTACAGTGGAGTATGGAGCATGTCCTGGTGGAGTGGAGTGGACTACCAGTGGAACAACTCC    IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	237 1980 247 2070 2160 273 2250 303
GCTKGe GCTKCD GCTKCD GCTKGe GCTKGe GCTKCD	$\label{eq:response} ArgPheSerIleThrProSerThrAsnThrGlnSerSerSerLeuVallleSerGlyLeuArgTyrSerAspAlaGlyAspTyrMetCysThrValLysTyrGlyAlaCysProGlyGlyValAspCysSerGlyThrThrPrTGGTCTCAGATATTCTGATGCAGGAGACTACATGTGTACAGTGGAGTATGGAGCATGTCTGGGTGGAGTGGACTGCAGTGGAACTACCCCInic and an and an antice and an antice and antice an$	237 1980 247 2070 2160 273 2250 303 2340
GCTKGe GCTKCD GCTKCD GCTKGe GCTKGe GCTKCD GCTKCD	ArgPheSerIleThrProSerThrAsnThrGInSerSerSerLeuValleSe    rGlyLeuArgTyrSerAspAlaGlyAspTyrMetCysThrValLysTyrGlyAlaCysProGLyGlyValAspCysSerGlyThrThrPr    TGGGTCTCAGATATTCTGATGCAGGAGACTACATGTGTACAGTGAGTG	237 1980 247 2070 2160 273 2250 303 2340
GCTKGe GCTKCD GCTKCD GCTKGe GCTKGe GCTKCD GCTKCE	ArgPheSerIleThrProSerThrAsnThrGlnSerSerSerLeuValleSe    rGlyLeuArgTyrSerAspAlaGlyAspTyrMetCySThrValLysTyrGlyAlaCysProGlyGlyValAspCysSerGlyThrThrPr    TGGTCTCAGATATTCTGATGCAGGAGACTACATGGTGTACAGTGGAGTAGGAGCATGTCGGAGTGGACTGCAGTGGACTACAGTGGAACAACTCC	237 1980 247 2070 2160 273 2250 303 2340
GCTKGe GCTKCD GCTKCD GCTKGe GCTKCD GCTKCD GCTKCE	ArgPheSerIleThrProSerThrAsnThrGlnSerSerSerLeuVallleSe    rGlyLeuArgTyrSerAspAlaGlyAspTyrMetCysThrValLysTyrGlyAlaCysProGlyGlyGlyValAspCysSerGlyThrThrPr    TGGTCTCAGATATTCTGATGCAGGAGACTACATGTGTACAGTGAAGTATGGAGCATGTCCTGGTGGAGTGGACTGCAGTGGAACAACTCC	237 1980 247 2070 2160 273 2250 303 2340
GCTKGe GCTKCD GCTKCC GCTKGe GCTKGe GCTKCD GCTKGe GCTKCD	ArgPheSerIleThrProSerThrAsnThrGlnSerSerSerLeuValleSe    rGlyLeuArgTyrSerAspAlaGlyAspTyrMetCysThrValLysTyrGlyAlaCysProGlyGlyGlyValAspCysSerGlyThrThrPr    TGGTCTCAGATATTCTGATGCAGGAGACTACATGTGTACAGTGGAGTGGAGCAGGCCTGGTGGAGTGGACTGCAGGTGGAACAACTCC	237 1980 247 2070 2160 273 2250 303 2340
GCTKGe GCTKCD GCTKCD GCTKGe GCTKGe GCTKCD GCTKGe GCTKCD	ArgPheSerIleThrProSerThrAsnThrGlnSerSerSerLeuVallleSe    rGlyLeuArgTyrSerAspAlaGlyAspTyrMetCysThrValLysTyrGlyAlaCysProGlyGlyValAspCysSerGlyThrThrPr TGGTCTCAGATATTCTGATGCAGGAGACTACATGTGTACAGTGGAGTATGGAGCATGTCCTGGTGGAGTGGACTGCAGTGGAACAACTCC    i/i/i/i/i/i/i/i/i/i/i/i/i/i/i/i/i/i/i/	237 1980 247 2070 2160 273 2250 303 2340 333
GCTKGe GCTKCD GCTKCD GCTKGe GCTKGe GCTKCD GCTKGe GCTKCD	ArgPheSerIleThrProSerThrAsnThrGInSerSerSerLeuVallleSe    rGlyLeuArgTyrSerAspAlaGlyAspTyrMetCysThrValLysTyrGlyAlaCysProGlyGlyValAspCysSerGlyThrThrPr    TGGTCTCAGATATCTGATGCAGGAGACTACATGTGTACAGTGGAGTGGAGTAGGAGCTATCGTGGGGAGTGGACTGCAGTGGAACAACTCC    IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	237 1980 247 2070 2160 273 2250 303 2340 333 2340
GCTKGe GCTKCD GCTKCD GCTKGe GCTKCD GCTKGe GCTKCD GCTKGe	ArgPheSerIleThrProSerThrAsnThrGInSerSerSerLeuVallleSe    rGlyLeuArgTyrSerAspAlaGlyAspTyrMetCysThrValLysTyrGlyAlaCysProGlyGlyValAspCysSerGlyThrThrPr TGGTCTCAGATATTCTGATGCAGGAGACTACATGTGTACAGTGGAGTATGGAGCATGTCCTGGTGGAGTGGACTGCAGTGGACAACTCC	237 1980 247 2070 2160 273 2250 303 2340 333 2340
GCTKGe GCTKCD GCTKCD GCTKGe GCTKCD GCTKCE GCTKCD GCTKGE GCTKCE	ArgPheSerIleThrProSerThrAsnThrGlnSerSerSerLeuValIleSe    rGlyLeuArgTyrSerAspAlaGlyAspTyrMetCysThrValLysTyrGlyAlaCysProGlyGlyClyValAspCysSerGlyThrThrPr TGGTCTCAGATATTCTGATGCAGGAGACTACATGTGTACAGTGGAGTAGGAGCATGTCCTGGTGGAGTGGACTGCAGTGGAACAACTCC	237 1980 247 2070 2160 273 2250 303 2340 333 2340
GCTKGe GCTKCD GCTKCD GCTKGe GCTKCD GCTKGe GCTKCD GCTKGE GCTKCD	$\label{eq:relevant} ArgPheSerIleThrProSerThrAsnThrGlnSerSerSerLeuValIleSe rGlyLeuArgTyrSerAspAlaGlyAspTyrMetCysThrValLysTyrGlyAlaCysProGlyGlyGlyValAspCysSerGlyThrThrPrTGGTCTCAGATATTCTGATGAGGGAGGTACATGTGTCAGATGGAGGAGGTACATGTGTACAGTGGAGGGAG$	237 1980 247 2070 2160 273 2250 303 2340 333 2430
GCTKGe GCTKCD GCTKCD GCTKGe GCTKCD GCTKCD GCTKCD GCTKCD	$\label{eq:light} ArgPheSerIleThrProSerThrAsnThrGlnSerSerSerLeuValIleSe rGlyLeuArgTyrSerAspAlaGlyAspTyrMetCysThrValLysTyrGlyAlaCysProGlyGlyValAspCysSerGlyThrThrPr TGGTCTCAGATATTCTGAGGGGGGACTACATGTGTACAGGGGGGGG$	237 1980 247 2070 2160 273 2250 303 2340 333 2430 333
GCTKGe GCTKCD GCTKGe GCTKGe GCTKGe GCTKCD GCTKGe GCTKCD GCTKGe GCTKCD	$\label{eq:lightproduct} ArgPheSerIleThrProSerThrAsnThrGlnSerSerSerLeuValIleSe rGlyLeuArgTyrSerAspAlaGlyAspTyrMetCysThrValLysTyrGlyAlaCysProGlyGlyValAspCysSerGlyThrThrPr TGGTCTCAGATATTCTGAGGGGGAGCACACATGCTCAGGGGGAGTATCCAGGGGGGAGCACACATGCC rGlyLeuArgTyrProAspAlaGlyAspTyrMetCysThrValLysTyrGlyProCysProAspGlyValAspCysSerGlyThrThrPr IGAT_GGCACAGGGCACACACTGCC rGlyLeuArgTyrProAspAlaGlyAspTyrMetCysThrValLysTyrGlyProCysProAspGlyValAspCysSerGlyThrThrPr IGA$	237 1980 247 2070 2160 273 2250 303 2340 333 2430 3350 2520
GCTKGe GCTKCD GCTKGe GCTKGe GCTKGe GCTKGe GCTKGe GCTKGe GCTKGE GCTKGE	$\label{eq:lighter} ArgPheserIleThrProSerThrAsnThrGlnSerSerSerLeuValIleSe} \\ rGlyLeuArgTyrSerAspAlaGlyAspTyrMetCysThrValLysTyrGlyAlaCysProGlyGlyValAspCysSerGlyThrThrPr TGGTCTCAGATATTCTGATGCAGGAGACTACATGTGTACAGTGGAGATATGGAGCATGTCGTGGGAGGAGCTGCAGTGGACAACATCC rGlyLeuArgTyrProAspAlaGlyAspTyrMetCysThrValLysTyrGlyProCysProAspGlyValAspCysSerGlyThrThrPr Ig 1 ) \\ oValThrGlyAspIleRisteuGluLeuProLit init con 1 \\ AGTCACTGGCAACATACATGTGAACATCCCATGTGTACGTGTACACTTAGCCAGTAGTCTTAGCCAGTGGATACAGGTTGCTAGGTTGCAGGTGGATGCCAGGTGGATGCCAGGTGGATGCCAGGTGGATGCCAGGTGGATGCCAGGTGGAACAACACCCC [ Ig 2 ] \\ aftroflyValIleRisteuGluLeuProLit init con 2 \\ [ Ig 2 ] \\ intron 1 : :::::::::::::::::::::::::::::::::$	237 1980 247 2070 2160 273 2250 303 2340 333 2430 333 2430 350 2520
GCTKGe GCTKGe GCTKGe GCTKGe GCTKGe GCTKGe GCTKGe GCTKGe GCTKGE GCTKGE	ArgPheSerIleThrProSerThrAsnThrGlnSerSerSerLeuValIleSe    rGlyLeuArgTyr <u>Ser</u> AspAlaGlyAspTyrMetCysThrValLysTyrGly <u>Ala</u> CysPro <u>Gly</u> GlyValAspCysSerGlyThrThrPr    rGGTCCCAGATATCTCGAGCAGGAGACTACATGGTGACAGGGGAGTAGGAGCAGCTGCCGGGGGGGAGCAGCACACTCCC    rGGTCCCAGATATCTCGAGCAGGAGACTACATGGTGACAGGGGAGAGAGCAGCAGCAGCGCCGGGGGGAGCAGCA	237 1980 247 2070 2160 273 2250 303 2340 333 2430 333 2430 350 2520

Fig. 1. See full caption below.

GCTKGe GCTKGe GCTKGe GCTKGe GCTKGe	GATATACCTGCTAGCTATGTACCAATGCATAAGGATGCATAAGAGTCAGAGGGACACTTCAAACCTTTAGGGTAGTGTGTGT	2610 2700 2790 2880 2970
GCTKGe GCTKcD	intron 2  ## lyGlySerAsnSerSerGlySerAsnSerGlyValleAlaGlyValLeuIleThrLeuLeuLeuLeuIle    TGTGTATCCACATCTGTAGGTGGCTCCCAACTCAAGTGGCAGTGACAGCGGAGTGATTGCTGGAGTTTTAATAACTCTTTTATTACTCATA	37 <u>4</u> 3060
GCTKGe GCTKcD	$\frac{TM}{I} \frac{JM}{J} \frac$	404 3150
GCTKGe GCTKcD	CysGlySerCysSerCysValProLeuLeuAlaAlaLeuLysGlyValLysLeuProThrArgHisArgGluAsnLeuAsnLysAsnGly TGTGGCTCATGTTCGTGTGTGCCCTCCCTTGCTGCACTGAAAGGTGTCAAACTCCCAACAAGACACGGAGAAAACTTGAACAAGAACGGA 	434 3240
GCTKGe GCTKcD	ThrArgLeuArgLeuAsnGluArgAsnHisIleAlaAspThrAsnThrGluIleTyrSerValValGlnLysProLeuLysLysIleAsn ACAAGACTGAGACTGAACGAGAGGAATCATATCGCAGACACCAATACTGAGATTTACAGTGTCGTACAGAAACCACTCAAGAAAATCAAC 	464 3330
GCTKGe GCTKcD	LysSerProProProLeuProProLeuThrLeuThrGluThrGluLeuAsnGluLeu <u>Met</u> SerIleAspGluLysGluGluLeuSerPro AAATCCCCACCACCACTTCCTCCCCCTCACACAGAGACTGAGTGAATGAA	494 3420
GCTKGe GCTKcD	IleGlnGluLysProThrArgArgAsnThrGlyLeuSerThrTyrSerGlnSerGlyThrIleProLysLeuAlaLysLeuThrLysLeu ATCCAAGAGAAACCGACACGAAGAAACACTGGTCTCTCGACCTACTCGCAGTCAGGGACCATCCCGAAACTGGCAAAGCTGACCAAACTG 	524 3510
GCTKG <del>e</del> GCTKcD	ArgLysPheLysMetLysGluAsnProIleTyrGlnSer <u>Val</u> Asp <u>Val</u> Leu <u>Val</u> LeuGluLeuGluLeuGlnValAspAsnThrLeuTyr    AGGAAGTTCAAGATGAAGGAGAACCCTATCTATCAGTCGGTGGAGCTGGAGCTGGAGCTGCAGGTGAAGACAACACCACTCTAC	554 3600
GCTKGe GCTKCD	AlaLeuProLeuLysProAsnSerThrArgAsnSerAlaSerPheThrAspAspLeuAlaSerAspProIleTyrSerValAlaIleAsn GCCCTCCCGTTAAAAACCGAACTCGAACAGTGCATCCTTCACCGATGACTTGGCATCTGACCCCATCTACAGCGTGGCGATAAAT 	584 3690
GCTKGe GCTKcD	ProSerMetPheThrLysArgSerSerThrIleGlyAsnAspAspAspLeuHisProTyrGlyProIleTyrAlaArgProIleLysGln CCAAGTATGTTCACCAAGAGGTCAAGCACCATTGGCAATGACGATGATCATCACCCCATACGGCCCCATCTACGCCAGACCTATCAAGCAG 	614 3780
GCTKGe GCTKcD	<u>JM</u> ][ <u>TK</u> LysMetArgGlnProLeuAsnValSerValAspAsnIleArgGluValLysGlnIleGlyValGlyGlnPheGlyAlaValValLeuAla    AAAATGAGGCAGCCCCTGAATGTTAGTGTGGATAACATCCGGGAGGTGAAACAGATTGGCGTCGGTCAGTTTGGAGCCGTGGTTCTTGCT    IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	644 3870
GCTKGe GCTKcD	GluMetThrGlyLeuSerGlySerAsnValAlaSerLeuProLysGlySerMetAsnAlaAspGlyValAlaLeuValAlaValLysLys GAAATGACTGGGCTCTCCGGGTCAAACGTTGCGTCCCTACCAAAAGGATCCATGAATGCTGACGGAGTAGCACTAGTGGCTGTGAAGAAA 	674 3960
GCTKGe GCTKcD	LeuLysProAspValSerAspGluVal <u>Arq</u> GlnSerPheAspLysGluIleLysPheMetSerGlnLeuGlnHisAspSerIleValGln CTARARCCGGATGTGAGTGACGAAGTACGGCAGTCGTTTGATAAGGAGATARAGTTCATGTCACGGCTCCAGCATGATAGCATCGTGCAG 	704 4050

GCTKGe GCTKcD	LeuLeuAlaValCysThrHisSerLysHisProPheIleValMetGluTyrMetGluAsnGlyAspLeuAsnGlnPheLeuGlnLysTyr CTTCTGGCCGTATGTACACACAGTAAGCATCCCTTCATTGTGATGGAATACATGGAGAATGGAGACCTCAACCAGTTCCTGCAGAAATAC 	734 4140
GCTKGe GCTKcD	Calculation of the second state of	764 4230
GCTKGe GCTKcD	ValTyr <u>LeuSerSer</u> LeuAsnTyrValHisArgAspLeuAlaThrArgAsnCysLeuValGlySerAsnPheArgIleLysIleSerAsp GTGTACCTCTCCTCGGCTCAACTACGTCCACCGAGACCTGGCCACAGAAACTGTCTCGTCGGTTCCAACTTCCGTATCAAGATCTCCGAC 	794 4320
GCTKGe GCTKcD	PheGlyMetSerArgAsnLeuTyrGluArgValTyrTyrArgValArgGlyArgAlaMetLeuProIleArgTrpMetAlaThrGluSer TTTGGCATGAGCCGCAACCTCTACGAGCGAGTCTACTACCGCGTCCGAGGCCGGGCAATGCTCCCCATTCGTTGGATGGCCACTGAGAGC 	824 4410
GCTKGe GCTKcD	PheTyrGlyArgPheSerGluLysSerAspAlaTrpAlaTyrGlyValThrValTrpGluIleTyrThrLeuGlyLysLysGlnProTyr TTCTATGGACGATTCTCAGAGAAGTCAGATGCGTGGGGGGGG	854 4500
GCTKGe GCTKcD	GluGluLeuAspAspGlnHisMetIleGlnAspAlaIleArgGlyThrGlyArgArgIleMetGlyArgPro <u>GluGlyCysProGlnAla</u> GAAGAGCTTGATGATCAGCATATGATCCAAGATGCTATTCGAAGGCACGGGTCGTAGGATCATGGGTCGACCCGAGGGGGTGTCCGCAGGCT 	884 4590
GCTKGe GCTKcD	ValTyrGluValLeuLeuArgCysTrpGluTyrAlaAlaAlaAspArgAlaThrPheLysGluIleHisAspSerLeuAsnLeuIleGln      GTGTATGAGGTGCTACTGCGGTGCGGGGGGTATGCTGCCGCAGATCGAGCCACGTTCAAAGAGATACATGACAGCCTGAACCTCATTCAA	914 4680
GCTKGe GCTKcD	<u>TK</u> ]    LeuAsnSer +    CTCAATAGCTGATGACTGTTTCACTTTCATTCAAGTTTTATAATGTATTTGTCGTCTGTCT	917 4769
GCTKGe GCTKcD	GTAAGTCAA <u>AATAAC</u> AGGTTTTTCACACTGCTGCATTGTTGGTTTTTACTATTGCGCATGCGTTGTTTTTACACGCGCAAAATAGTCAGG 	4859

GCTKGe TGGCGCTGCTCT

**Fig. 1.** Comparison of the nt as well as the deduced as sequence (three-letter code) both from the RTK gene, *GCTK*Ge, and from the RTK cDNA, *GCTK*cD. Identical nt between the two sequences are indicated (|); the nonidentical as are *underlined*. The donor and acceptor splice sites are also marked ([‡]); the location of the introns is indicated. The nt as well as the deduced as sequence of the *GCTK*Ge is

even more (Hunter 1987). This number is close to our present knowledge. It can be expected that many of these enzymes, especially if they are members of the same (or closely related) subfamilies, very probably have only a recent biological history. PTKs are divided into two major groups, the receptor tyrosine kinases (RTKs), which are membrane-spanning molecules with similar overall structural topologies, and the nonreceptor TKs, also composed of structurally similar molecules. Members of the PTKs are further classified on the basis of their structural (functional) similarities and divided into over 20 subfamilies (Hardie and Hanks 1995). Phylogenetic *numbered;* the putative termination of the deduced as sequence is marked (+). The potential transcription start site and the polyadenylation site are *underlined*. The CArG box-like sequence is *double underlined*. The different domains, the first and second Ig-like domain (Ig 1 and Ig 2), the transmembrane domain (TM), the juxtamembrane region (JM), and the TK-domain (TK), are indicated.

analysis, based on the alignment of TK-domain amino acid (aa) sequences, indicates that all TK proteins have one common ancestor (Hardie and Hanks 1995). Since PTKs were found only in the metazoan (multicellular) organisms it is reasonable to postulate that these enzymes derive from one common ancestral molecule (Müller 1995).

Sponges (Porifera) are the most primitive multicellular animal phylum; they are known to have existed since the Proterozoic period (Orlov 1971). Recent analysis of phylogenetically conserved proteins from the marine sponge *Geodia cydonium* (Porifera, Demospongiae,

#### 4871

kb 2.8 - - 3.3 1.6 -1.0 -0.6 -0.4 -0.3 -

**Fig. 2.** Northern blot analysis of *G. cydonium transcript* for RTK using the homologous cDNA as a probe. Further details are given under Materials and Methods.

Geodiidae), a lectin (Pfeifer et al. 1993a), the RTK (Schäcke et al. 1994c), and a homeodomain-like polypeptide (Kruse et al. 1994) strongly supports the assumption of the very old lineage of the sponge taxa and indicates that sponges branched off from the common metazoan tree over 600 million years ago (Müller 1995). Furthermore, the existence of conserved proteins, structurally and functionally similar to higher metazoan proteins, speaks in favor of a monophyletic origin of Porifera and other "classical" metazoan animals (Müller 1995). Fossil species belonging to the genus Geodia, Lamarck 1815, have Cretaceous (145-75 million years ago)-to-recent range (De Laubenfels 1955); G. cydonium is a long-living and large sponge species (up to 70 cm in diameter) which predominantly reproduces asexually (Arndt 1930).

Here we show that—in contrast to similar sequences from higher metazoa—the RTK gene from *G. cydonium* is not interspersed by introns.

#### **Materials and Methods**

*Materials.* Enzymes for recombinant DNA techniques and vectors were obtained from Stratagene (Heidelberg, Germany), Boehringer (Mannheim, Germany), Promega (Madison, WI, USA), Epicentre Technologies (Madison, WI, USA), and USB (Cleveland, OH, USA).

*Sponge*. The specimens of *Geodia cydonium* (Porifera, Demospongiae, Tetractinomorpha, Astrophorida, Geodiidae) were collected near Rovinj (Croatia). The material was immediately frozen in liquid nitrogen until use.

Preparation of G. cydonium Genomic Library. Genomic DNA from G. cydonium was isolated from frozen tissue using ultracentrifugation in guanidine isothiocyanate/cesium chloride gradient and purified according to standard phenol-chloroform extraction procedure (Ausubel et al. 1995). The small-scale reactions of partial restriction endonucle-ase digestion of G. cydonium genomic DNA were performed using Sau3A I to generate fragments of 15–23 kb. With the cos-sites pro-



**Fig. 3.** Gene structure of *GCTK*Ge, coding for an RTK from *G. cydonium.* The two introns are inserted (1) between the first and the second Ig-like domain (160 bp) and (2) between the Ig-like domain and the transmembrane domain (508 bp). The other parts of the sponge RTK, the transmembrane domain (TM, 84 bp), the juxtamembrane region (JM, 693 bp), and the TK-domain (TK, 879 bp), form one exon.

tected by ligation,  $\lambda$ GEM 12 vector (Promega) was digested with *XhoI*. Both  $\lambda$ GEM 12/*XhoI* and *G. cydonium* genomic DNA/*Sau3*A I were partially filled-in using the Klenow Partial Fill-In Kit (Stratagene). Since the partial fill-in procedure prevents the self-ligation reactions of vector arms, central stuffer, and genomic fragments, the primary ligation of genomic DNA into  $\lambda$ GEM 12 arms was performed at 4°C overnight. Optimal packaging efficiency of ligated DNA products was performed using Max Plax Packaging Extract (Epicentre).

Isolation of G. cydonium Genomic Clones for Receptor Tyrosine Kinase. Southern hybridization of  $10^6$  pfu of G. cydonium genomic library was performed under moderately stringent conditions (42°C, 50% formamide) with plasmid probe containing G. cydonium RTK cDNA (termed *RTK*cD; accession No. X77528; Schäcke et al. 1994c) applying the random priming labeling DIG system from Boehringer (Ausubel et al. 1995). Subsequently filters were washed with 2× SSC, 1× SSC, 0.5× SSC, and 0.25× SSC at room temperature and finally with 0.1× SSC at 55°C (all solutions contained 0.1% NaDodSO₄). Hybridization signals were detected with anti-DIG alkaline phosphatase, Fab fragment (Boehringer). Several genomic clones containing coding sequence for the G. cydonium of RTK were isolated. Restriction endonuclease mapping of plaque-purified positive isolates was performed by Southern hybridization (Ausubel et al. 1995) using DIG-labeled cDNA probe.

DNA Sequence Analysis. Genomic clone fragments of interest were subcloned into pBluescript vector for sequencing and further restriction analysis. Double-stranded DNA sequencing was performed according to the dideoxy chain termination method of Sanger (Sanger et al. 1977) with  $[\alpha^{35}S]$ dATP using the DNA Sequencing kit (USB). Ambiguities in exon parts were resolved using the IsoTherm DNA Sequencing kit (Epicentre). Sequences were composed and analyzed using the computer program PCGene (1995).

Northern Blot. Sponge tissue was frozen in liquid nitrogen and then pulverized. RNA was then extracted from the tissue using TRIzol Reagent (GibcoBRL) as recommended by the manufacturer, with an additional isolation step for samples with high polysaccharide content. Poly(A) RNA was purified with Oligotex mRNA kit (Qiagen). One microgram of mRNA was electrophoresed through 1% formaldehyde/ agarose gel and blotted onto Hybond N⁺ membrane following the instructions of the manufacturer (Amersham). The RNA molecular weight marker II (Boehringer) was used for size estimates.

Hybridization was performed with the cDNA of *RTK*cD at 65°C overnight in the following buffer: 0.25 M NaH₂PO₄ (pH 7.2), 1 mM EDTA, 5% SDS, and 0.5% blocking reagent (DuPont). Washes were done at 65°C as follows: twice in 0.125 M NaH₂PO₄ (pH 7.2), 2.5% SDS, 0.05 mM EDTA, and then twice in 0.025 M NaH₂PO₄ (pH 7.2), 0.5% SDS, 0.01 mM EDTA (Ausubel et al. 1995). The film was exposed for five days at  $-80^{\circ}$ C.

246

	Sobbomain I (All binding Sicc)		-	week and the second sec		
				Vavk	1000	
GCTKGe	IREVKQIGVGQFGAVVLAEMTGLSGSNVASLPKGSMNAD	100%	GVAL	VAVKKLKPDVSDE	100%	
GCTKcD	***************************************	100%	****	*********	100%	
CAK_Human	*VLKWEL*E*A**K*F***CHNLLPEQ	41%	DKM*	****A**-EA*ES	53%	
TRKB_Mouse	*VLKREL*E*A**K*F***CYNLCPEQ	41%	DKI*	****T**-*A**N	60%	
TRKC_Pig	*VLKREL*E*A**K*F***CYNL*PTK	41%	VKM*	****A**-*PTLA	53%	
ABL_Dmela	*MMKHKL*G**Y*E*YE*VWKR	32%	YGNI	****T**E*TM	47%	
ABL_Human	*TMKHKL*G**Y*E*YEGVWKK	27%	YSLI	****T**E*TM	47%	
ROS Human	LTLRLLL*S*A**E*YEGTAVDIL*VGS	23%	*EIK	(****T**KGST*Q	478	
IG1R_Human	*TMSREL*Q*S**M*YEGVAK*VVKDE	27%	PETR	**I*TVNEAA*MR	27%	
INSR Human	*TLLREL*Q*S**M*YEGNARDIIKGE	27%	AETR	****TVNESA*LR	33%	
IRR Human	*SIIREL*Q*S**M*YEGLAR**EAGE	32%	ESTF	**L*TVNELA*PR	27%	
FGR1 Human	LVLGKPL*E*C**Q*****AI**DKDKPN	45%	RVTK	****M**S*ATEK	47%	
TEC Dmela	LMLMEEL*S****V*RRGKWR	27%	*SID	T***MM*EGTM	33%	
FER Human	VILGELL*K*N**E*YKGTLK	23%	DKTS	****TC*E*LPQ*	47%	
FES Human	LVLGE***R*N**E*FSGRLRA	32%	DNT*	****SCRETLPPD	33%	
SEV Dmela	I.KI.LRFL*S*A**E*YEGOLKTEDSE	23%	EPOR	**I*S*RKGA*	33%	
TIE Mouse	TKFODV**E*N**O*LKRIKKDG	27%	LRME	A*I*RM*EYA*KD	27%	
III_Mouse						
0007/0-	UDOGEDVETVENCOL 100% OUDSTUOLLAUCTUS 1	008				
GCTKGE	VKQSFDKEIKFMSQL- 100% QHDSIVQLLAVCINS I	000				
GCTKCD	$\mathbf{F}_{\mathbf{v}}^{\mathbf{v}} \mathbf{v}_{\mathbf{v}}^{\mathbf{v}} \mathbf{v}_{\mathbf{v}}^{\mathbf{v}}} \mathbf{v}_{\mathbf{v}}^{\mathbf{v}} \mathbf{v}_{\mathbf{v}$	738 778				
CAK_Human	A*ND*L**V*I**K*= 53% KDPN*IK**G**VQD	338				
TRKA_Human	A**D*QK*AELLIM*- 33% **QH**KFFG***EG	4/6				
TRKB_Mouse	A*KD*HR*AELLTN*- 2/8 **EH**KFYG**VEG	408				
TRKC_Pig	A*KD*QK*AELLTN*- 2/8 **EH**KFIG**GDG	406				
ABL_Dmela	ALKD*LE*AAI*KEM- 20% K*PNL***IG***KE	4/6				
ABL_Human	EVEE L AAV AEI - 2/8 A PAL A G A A	226				
ROS_Human	EKIEALAAHLAAKF- 336 MANALKYAGAALKA	223 278				
IGIR_Human	$E^* I E^* L N^* A S V^* R E F = 278 N C H V * R^* A * V S V S K C + 1 + 1 + 1 + 1 + 2 + 2 + 2 + 3 + 2 + 3 + 3 + 3 + 3 + 3$	278				
TRB_Human	ETENINASVARGE 27% ICHNVARAGAVSKG	278				
FRC1 Numan	DI CDI IC*WEW*KWIG 139 K*KN*IN**GA**OD	40%				
TEC Drola	GEDDATEAAYUATKA 339 OHDNI.***VG***KH	40%				
IEC_DHEIA	$\frac{3600}{16} + \frac{3}{16} + \frac{3}{1$	479				
FER_Human	$\frac{1}{1} \frac{1}{1} \frac{1}$	478				
FES_Human	FEAFILO*ACI **NE_ 20% S*FM K 10 QK	338				
TIF Mouso	DUED + AGAI FUI CK+G 208 H*DN*IN**GA*E*R	40%				
IIE_Mouse	DRFR AG LAVIER G 20% IF TH TH OR I K	100				
	CURDOWATN V			SIT	BDOMATN VTA	
COMPCA	VUDETUNEVNENCDI NOFI OKVONUDDDSALVSNO		1008	TPPSTLLV	MAVOTASCHVYLSSI.	100%
GCIRGE	KHFF1VMEIMENGDLNQFLQKIQMVDDDSALISNQ		100%	********	***********	87%
GOIRCD		OGPT	50%	*SVDM**H	 V*&**************	57%
CAR_Human	DEL INTETTOTION AND AND AND AND AND AND AND AND AND AN		38%	LGLGO**A	V*S*V*A*******	48%
TRRA_Human	RFLLM*F***KU********************************		388	LTO*0W*H	T*O***A*****A*O	52%
TRKB_MOUSE	DPLIM*F***KH********************************	F	388	LGL*0W*H	T*S**C*******	52%
ARKC-PIG	DDEX+IM+E+CU+N+ID++BCACEEM		279		********************	618
ABL_Dmeia			278	WNAWV***	*************	48%
ABL_Human	PPFI*11*F*11****************************		209	TTTVD*VD	LC*D*SK*C***ERM	308
KOS_Human	CPUI · IL · L · · · · · · L II · K · AK · AIF IGF ·		279	PSL*KMIO	**GE**D**A**NAN	398
IGIR_Human	QPTLVI **L*IK***KSI*KOLKFEMEMAFVLAF		219	PUL RMIQ	**************************************	358
INSK_Human	QPTLV * * * L * AR * * * KSL * KSL KPEAEAAFGKFF		279	PILQEMIQ	**GF**D**A**AAN	35%
IKK_Human	QPTLVI**L*TK***K5H*K5LKPEAENNPGLPQ		210	FALGEAIQ	C*V*V*P**F**A*K	308
FGRI_Human	GFLIVIV ~ AGA ~ N ~ KEI ~ AKKFFGLEICINPSANPEE	×	218	GNMGI**D	*CT*VSK******	359
TEC_DMela	ЛГІІ""Т°°°ЛП°Э°LЯІ"ККПЕКТЦІ		279	TRIKO**D	FSI.DA*A**I.**F*F	30%
FER-numan	VD1A+++1 NOG++L1 #++DAECyB		218	T.BARK***U	*VGDA*A**E**E*S	438
FES_RUMAN	ELLE TARRAY A ***I CANDA DDALCUULDUUUUU		319	LSI.*E**A	*CIDV*N*CS**EDM	35%
TTE Mouse	CVI.VI.AT**ADH*N*I.D**D*SDM.FT*D*FATAN		278	STA**LSSOO**H	F*ADV*R**D***OK	48%
TTO MOUSE	GININGI GRANN AND CONVERT		- 1 0	TOPAA	21	

SUBDOMATN TT

SUBDOWATH T (ATP binding site)

Fig. 4. Multiple aa alignment of G. cydonium TK-domain (one-letter code), both of the deduced aa sequences from the gene (GCTKGe) and from the cDNA (GCTKcD) with 16 TK-domains of other PTKs displaying highest homology. The following sequences are shown: cell adhesion kinase (CAK Hu; accession No. L20817; Perez et al. 1994); nerve growth factor receptor (TRKA_Hu; P04629; Martin-Zanca et al. 1986); neurotrophin-4 receptor (TRKB_Mouse; P15209; Klein et al. 1989); NT-3 growth factor precursor (TRKC_Pig; P24786; Lamballe et al. 1991); DASH/ABL proto-oncogene tyrosine kinase from Drosophila melanogaster (ABL_Dmela; P00522; Henkemeyer et al. 1988); ABL proto-oncogene tyrosine kinase (ABL_Human; P00519; Shtivelman et al. 1986); ROS proto-oncogene tyosine kinase (ROS_Human; M35106; Birchmeier et al. 1990); insulin-like growth factor 1 receptor precursor(IG1R_Human; P08069; Ullrich et al. 1986); insulin receptor

precursor(INSR_Human; P06213; Ullrich et al. 1985); insulin receptorrelated receptor (IRR_Human; P14616; Shier and Watt 1989); basic fibroblast growth factor receptor 1 precursor (FGR1_Human; P11362/ P17049); Drosophila melanogaster SRC protein tyrosine kinase (TEC Dmela.; M11917); FES/FPS-related PTK (FER Human; J03358; Hao et al. 1989); FES/FPS protein-tyrosine kinase (FES_Human; P07332; Roebroek et al. 1985); Drosophila melanogaster sevenless receptor PTK (SEV_D. mela.; P13368; Bowtell et al. 1988); TIE protein-tyrosine kinase (TIE_Mouse; X71425; Sato et al. 1993). The delineation of the TK subdomains I-XI is adopted from Hardie and Hanks (1995). The homologies of the subdomains with respect to the sponge aa sequence from GCTKGe are given in percent. The consensus sequences mentioned in the text are shown.

35%

35%

48%

su	BDOMAIN VIB (CATALYTIC LOOP)	SUBDOMAIN VII		(
	HRDLATRN	DFG Y YY		
GCTKGe	NYVHRDLATRNCLVGSN 100%	FRIKISDFGMSRNLYERVYYR 1	00%	
GCTKcD	***************************************	********* <b>K</b> *********	95%	
CAK_Human	*F************************************	*T***A********************************	76%	
TRKA_Human	HF**********QG 76%	LVV**G*****DI*STD***	57%	
TRKB_Mouse	HF************E* 82%	LLV**G*****DV*STD***	57%	
TRKC_Pig	HF************************************	LLV**G*****DV*STD***	57%	
ABL_Dmela	**I*****A****D* 82%	KLV*VA***LA*LMRDDT*T-	29%	
ABL_Human	*FI*****A*****E* 76%	HLV*VA***L**LMTGDT*T-	33%	
ROS_Human	HFI****A****SVKDYTSP 59%	RIV**G***LA*DI*KND***	48%	
IG1R_Human	NK*****A***M*AED 65%	*TV**G****T*DI**TD***	62%	
INSR_Human	KF*****A***M*AHD 59%	*TV**G****T*DI**TD***	62%	
IRR_Human	KF*****A***M*SQD 59%	*TV**G****T*DV**TD***	62%	
FGR1_Human	KCI****A**V**TED 53%	NVM**A***LA*DIHHID**K	38%	
TEC_Dmela	**I*****A*****************************	NVV*VA***LA*YVLDDQ*T-	29%	
FER_Human	*CI*****A******************************	NVL********QEDGGV*S-	48%	
FES_Human	CCI*****A****TEK 59%	NVL********EEADG**AA	52%	
SEV_Dmela	HF*****C****TESTGSTDRR 65%	RTV**G***LA*DI*KSD***	48%	
TIE_Mouse	QFI*****A**I***E* 65%	NIA**A***L**GQE**V-	43%	
	SUBDOMAIN VIII	SUBDOMAIN IX		
	PIRW A E			
GCTKGe	VRGRAMLPIRWMATES- 100% FYGRFS	<b>EKSDAWAYGVTVWEIYTLGKKOPY</b>	EEL 100%	
GCTKCD	***************************************	****	*** 100%	
CAK Human	*0***V********************************	<b>FA**V**F***L**VLM*CRA**F</b>	GQ* 45%	
TRKA Human	*G**T*********************************	CE**V*SF**VL***F*Y*-***W	¥O* 48%	
TRKE Mouse	*G*HT**********************************	CE**V*SL**VL***F*Y*-***W	¥O* 48%	
TRKC Pig	*G*HT**********************************	TE**V*SF**IL***F*Y*-***W	FO* 48%	
ABL Dmela	AHAG*KF**K*T*P*GL 38% A*NK**	C***V**F**LL***A*Y*M S**	PAI 52%	
ABL Human	AHAG*KF**K*T*P**L 44% A*NK**	[***V**F**LL***A*Y*M S**	PGI 52%	
ROS Human	K**EGL**V****P**L 63% MD*1*T	TO**V*SF*ILI***L***H-***	PAH 45%	
TG1R Human	KG*KGL**V***SP**L 50% KD*V*T	TY**V*SF**VL***A**AE-***	OG* 48%	
TNSR Human	KG*KGL**V****P**L 56% KD*V*T	TS**M*SF**VL***TS*AE-***		
IRR Human	KG*KGL**V****P**L 56% KD*I*T	TH**V*SF**VL***V**AE-***	QG* 48%	
FGR1 Human	KTTNGR**VK***P*AL 38% FDRIYT	HO**V*SF**LL***F***G-S**	PGV 39%	
TEC Dmela	SS*GTKF**K*APP*VL 31% N*T***	S***V*****LM***F*C**-M**	GR* 64%	
FER Human	SS*LKOI**K*T*P*AL 38% N***Y*	SE**V*SF*ILL**TFS**V-C**	PGM 42%	
FES Human	SG*SROV*VK*T*P*AL 31% N***Y*	SE**V*SF*ILL**TFS**A-S**	PNL 42%	
SEV Dmela	KE*EGL**V***SP**L 50% VD*L*T	TO**V**F**LC***L***O-***	AAR 52%	
TIE Mouse	KKTMGR**V***I**L 50% N*SVYT	TN**V*S***LL***VS**G-T**	CGM 42%	
	SUBDOMAIN X	SUBDOMAI	N XI	
GCTKGe	DDQHMIQDAIRGTGRRIMGRPEG 1	00% CPQAVYEVLLRCWEYAAADR	ATFKEIHDSLNLIQLNS	100%
GCTKcD	***D**********************************	90% VAGC*-RGAT*********	*****	76%
CAK_Human	T*EOV*EN*GEFFRDQGROVYLS**PA	25% *** <b>AV</b> ** <b>LM</b> **** <b>SRESEQ</b> *	PP*SQL*RF*AEDA**T	41%
TRKA_Human	SNTEA*DCITQ*RELE**RA	20% **PE**AIMRG**QREPQQ*	HSI*DV*AR*QALAQAP	27 %
TRKB_Mouse	SNNEV*ECITQ*RVLQ**RT	20% ***E***LM*G**QREPHT*	KNI*S**TL*QNLAKA*	41%
TRKC_Pig	SNTEV*ECITQ*RVLE**RV	20% ***E**D*M*G**QREPQQ*	LNI***YKILHALGKAT	35%
ABL_Dmela	*LTDVYHKLDK*YRME**P*	25% **PE**DLMRQ**QWD*T**	P***S**HA*EHMFQE*	43%
ABL Human	*LSQVYELLEKDYRME****	25% **EK***LMRA**QWNPS**	PS*A***QAFETMFQE*	38%

IG1R_Human SNEQVLRFVME*GLLD-----K*DN 10% **DMLF*LMRM**Q*NPKM*PS*L**IS*IKEEMEPG 10% **ER*TDLMRM**QFNPKM*P**L**VNL*KDDLHP* INSR_Human SNEQVLKFVMD*GYLD-----Q*DN IRR Human SNEQVLKFVMD*GVLE----EL** 15% **LQLQ*LMS***QPNPRL*PS*TH*L**IQEELRP* FGR1_Human PVEELFKLLKE*HRMD-----K*SN 10% *TNEL*MMMRD**HAVPSQ*P***QLVED*DR*VALT 15% *AKET*D*MKL**SHGPEE*PA*RVLM*Q*A*VA---TEC Dmela KNTEVVERVQ**IILE----K*KS TN*QAREQVE**YRMS-----A*QH 20% **EDISKIMMK**D*KPEN*PK*S*LQKE*TI*KRKL FER Human SN*QTREFVEK*GRLP----C**L 20% **D**FRLMEQ**A*EPGQ*PS*ST*YQE*QS*RKRH FES Human 10% *TEKL*SL**L**RTDPWE*PS*RRCYNT*HA*STDL SEV_Dmela NNFEVLAHVKE*GRLQ-----Q*PM TIE Mouse TCAELYEKLPQ*YRLE-----K*LN 10% *DDE**DLMRQ**REKPYE*PS*AQ*LV***-----

SNLDVLNYVQT*GRLE-----P*RN 10% **DDLWNLMTQ**AQEPDQ*P**HR*Q*Q*Q*FRR*F

Fig. 4. Continued

**ROS Human** 

## **Results and Discussion**

# Cloning of the Ancient Receptor Tyrosine Kinase Gene in G. cydonium

We used the Geodia cydonium RTK cDNA probe (termed here GCTKcD; accession No. X77528) (Schäcke et al. 1994c) to screen this sponge's genomic DNA library (Sambrook et al. 1989). A genomic DNA clone containing the entire coding sequence of sponge RTK was analyzed. It contains the gene for RTK, designated here GCTKGe (accession No. X94128), which has a length of 4,871 bp (Fig. 1).

43%

32%

30%

328

32%

27%

27%

278

328

278

30%

A comparison of the corresponding segments of the nucleotide (nt) as well as the deduced amino acid (aa) sequences of GCTKcD and GCTKGe revealed that they differ only slightly 105 out of 2,181 nt and 32 out of 727 aa (Fig. 1).

The size of the sponge RTK mRNA was identified by Northern blotting using the cDNA of GCTKcD as a probe. An mRNA species of 3.3 kb was detected (Fig. 2). The deduced aa sequence of the exons of *GCTK*Ge reveals an  $M_r$  of 101,309—917 aa (corresponding to 2.8 kb).

## 5'- and 3'-Flanking Region of the GCTKGe Gene

The typical translation start site  $A/G_{-3}/G_{+4}$  (Kozak 1986), nt 1268–1274, is present. The 5'-nontranslated region contains potential promotor elements, TATA box (nt 646–651), GC-box (nt 623 as a center), and Cap signals (nt 669, 672, and 691 as centers) (PC/GENE 1995; program EUKPROM). In addition a CArG box-like sequence, CCTATATGG (nt 97–105), is present in the potential promotor region, which is known to be the target for the serum response factor, a transcription factor which is involved in the RTK signaling pathway (Shore and Sharrocks 1995). Recently, this cDNA was also cloned from *G. cydonium* (to be published).

A typical polyadenylation site, AAUAAA (Zarkower et al. 1986), is missing in both *GCTK*Ge and *GCTK*cD. This site is also absent in other cDNAs cloned from *G. cydonium*, e.g., ubiquitin (Pfeifer et al. 1993b; EMBL accession No. X70917) the lectin (Pfeifer et al. 1993a; No. X93925). In the 3'-nontranslated region (length of 112 nt) the sequence AAUAAC (Wickens and Stephenson 1984) which is present 23 nt upstream of the poly(A) sequence might function as a polyadenylation signal (Fig. 1). The stop codon UGA is located 111 nt from the poly(A) sequence.

# Gene Structure of Sponge RTK

The coding sequence of RTK contains three exons. The putative structure of the sponge RTK gene, GCTKGe, shows (1) the extracellular part, comprising two complete immunoglobulin (Ig)-like domains (Schäcke et al. 1994a, b) (first aa 106-247 and second aa 248-350 with respect to the aa in the sequence deduced from GCTKGe), (2) the transmembrane domain (aa 366–392), (3) the juxtamembrane region (aa 393–626), and (4) the catalytic TK-domain (aa 627-917) (Fig. 3). Two introns have been found. The first is located between the two Ig-like domains (nt 2013–2172) and the second intron between the second Ig-like domain and the transmembrane region (nt 2482-2989) in the extracellular part of GCTKGe. The introns are of medium size (160 bp and 508 bp, respectively). The consensus sequence at the boundary of the first intron GT (donor) and AG (acceptor) is a typical one, while in the second intron the nt GC and AG are less frequently used (Stephens and Schneider 1992). The 247th aa (Leu) is interrupted by the first intron and the 350th aa (Gly) by the second intron.

However, the rest of the gene, comprising the trans-

membrane domain, the juxtamembrane region and the catalytic TK-domain is coded by one single exon.

#### Tyrosine Kinase Catalytic Domain

A search for homology of *G. cydonium* TK-domain aa sequence deduced from the gene sequence was performed (PC/GENE 1995). All 50 most homologous sequences to sponge RTK contained the TK-domains of PTKs. Many of those PTKs were orthologous gene products from different organisms. One representative specimen from each group of closely related enzymes was used for the multiple sequence alignment. The alignment was computed (CLUSTAL; PC/GENE 1995) and subsequently manually adjusted to delineate the TK subdomains I–XI (aa 535–825), as proposed by Hardie and Hanks (1995) (Fig. 4).

The sponge TK-domain contains all conserved aa, or streches of aa, known to be important for the function of these enzymes (Geer et al. 1994) (Fig. 4). In detail, subdomain I shows the ATP-binding site (consensus: GXGXXGXV) and subdomain II shows Lys in the consensus vavK which is required for kinase activity. This K (Lys) is likely to be involved in the phosphotransfer reaction, possibly functioning in proton transfer (Russo et al. 1985). The aa D (Asp) and N (Asn) in subdomain VIB as well as the DFG trimer in subdomain VII are present in the sponge sequence; DFG has been implicated in ATP binding (Hanks et al. 1988). This triplet is the most conserved portion in the catalytic domain and is surrounded for two positions on both ends by the hydrophobic or near-neutral residues I (I1e) and S (Ser) (5'position) and M (Met) and S (Ser) (3'-position). Asp (D) in subdomain VIB and Asp (D) in subdomain VII are thought to interact with the phosphate groups of ATP via a Mg²⁺ salt bridge (Brenner 1987). Y (Tyr) (aa 181 of RT domain; subdomain VII) undergoes phosphorylation (tyrosine kinase phosphorylation site) - located seven residues to the C-terminal side of an R (Arg); this Y (Tyr) is located adjacent to the Y (Tyr) (aa 182) which is the potential autophosphorylation site. Signatures in subdomains III, IV, V, VIA IX, X, and XI are less well conserved.

Highest homology in the *G. cydonium* TK-domain with 45% was found in the recently discovered human cell adhesion kinase CAK RTK (Johnson et al. 1993; Perez et al. 1994), a protein which is not closely related to other known RTKs. CAK RTK represents a new class of RTKs. Second-best homology (over 43%) was found within the subfamily of RTK with genes coding for the nerve growth factor receptors (*TRKB*Mo, *TRKCPi*, and *TRKA*Hu). These TKs also contain two Ig-like domains in the extracellular part of the proteins. RTK genes of the insulin receptor subfamily (*IRR*Hu, *IG1R*Hu, *INSR*Hu) and of related enzymes (*ROS*Hu, *SEVD*m) also show significant



Fig. 5. Dendrogram deduced from the multiple alignment produced with CLUSTAL (PC/GENE 1995) of the TK-domain of RTK gene of *G. cydonium* (GCTKGe) with those of ten receptor TKs [neurotrophin-4 receptor (*TRKB*_Mouse); NT-3 growth factor precursor (*TRKC*_Pig); nerve growth factor receptor (*TRKA*_Human); ROS proto-oncogene tyosine kinase (*ROS*_Human); *Drosophila melanogaster* sevenless receptor PTK (*SEV*_Dmela); insulin receptor-related receptor (*IRR*_Human); insulin-like growth factor 1 receptor precursor (*IG1R_Human*); insulin receptor precursor (*INSR_Human*); basic fibroblast growth fac-

homology in the deduced TK-domain to the sponge *GCTK*Ge (35–37.5%). Homologies in the range of 32% were found with the fibroblast growth factor subfamily of RTKs (*FGR1*Hu), having three Ig-like domains in the extracellular part, and the *TIE* receptor PTK (*TIE*Mo; two Ig-like domains). Interestingly, several nonreceptor TKs were selected among the 50 proteins having highest homology in the TK-domain with the sponge RTK. The TK-domains of nonreceptor TK genes from the Abl subfamily (*ABL*Dm, *ABL*Hu) show remarkable homology (41%) and other selected nonreceptor TKs (*TEC*Dm, *FER*Hu, and *FES*Hu) still show a homology of over 35%.

#### Phylogenetic Tree

The dendrogram of the alignment of TK-domains of 15 PTKs produced by the CLUSTAL program is shown in Fig. 5. CAK RTK was not used here, because of its uniqueness (1) in its extracellular sequence as well as (2) in its structural motifs (Perez et al. 1994). All RTKs used for the dendrogram fall in one branch of the tree, while nonreceptor TKs are grouped in a second one; sponge RTK is placed in a separate branch, which splits off first from the common tree of metazoan PTKs.

# Introns Early or Late?

Most of the information about the PTKs in different organisms was obtained from studies of the corresponding

tor receptor 1 precursor (*FGR1_Human*) and TIE protein-tyrosine kinase (*TIE_Mouse*)] and of five nonreceptor TKs [*DASH/ABL* protooncogene tyrosine kinase from *Drosophila melanogaster* (*ABL_D*mela.); ABL proto-oncogene tyrosine kinase (*ABL_Human*); *Drosophila melanogaster* SRC protein tyrosine kinase (*TEC_Dmela*); FES/FPS-related PTK (*FER_Human*); FES/FPS protein-tyrosine kinase (*FES_Human*)]. The accession numbers and the references are given in legend to Fig. 3.

cDNAs. During the last few years the organization of a growing number of genes encoding these proteins has been analyzed. To the best of our knowledge, all studied genes contain introns in their TK-domain. Some of the genes encoding RTKs are over 100 kb long with over 20 introns, as in the case of the insulin receptor subfamily of RTKs (Seino et al. 1989). The insulin receptor gene contains five introns in the TK-domain. Comparison of the exon structure of the TK-domains of this and three other TK genes (ROS, SRC, and ERBB2) revealed that the exon-intron organization of this region has not been well conserved during evolution (Seino et al. 1989). According to the exon theory of genes (Gilbert et al. 1986; Gilbert 1987), proposed by Seino et al. (1989), the putative ancestral TK-domain may have been assembled from 13 exons; consequently, introns must have been lost in a more or less random fashion from individual genes. Genes encoding human RTKs with three, five, or seven Ig-like domains and the TK insert within the catalytic domain were recently studied in detail (Agnès et al. 1994; Rousset et al. 1995). In these reports introns at conserved positions within the TK-domains of these genes were found. These data were used for the establishment of the phylogenetic relationships between three related subfamilies of RTKs, and it was proposed that all genes most probably evolved from the common ancestor already "in pieces" by successive duplications involving entire genes. However, the time when the first duplication might have occurred was not discussed.

The tyrosine kinase domain of the ancient enzyme,

the RTK from the marine sponge *G. cydonium*, is — as mentioned — encoded by one single exon. There are many reasons to assume that *G. cydonium* RTK is the most ancient RTK analyzed so far (Schäcke et al. 1994a—c) that branched off first from the common tree of metazoan PTKs (Fig. 5). PTKs are found only in metazoan (multicellular) organisms and *G. cydonium* belongs to the oldest and the most simple metazoan phylum, the Porifera.

#### Conclusion

One can only speculate about the structure of the common ancestor gene for all TK-domains of PTK, including the G. cydonium RTK. According to the exon theory of genes, also called the "intron early" view (Gilbert et al. 1986; Gilbert 1987; Darnell 1978; Doolittle 1978), this common ancestor gene was present already in pieces. If this assumption should be correct, then (at least) in the sponge G. cydonium all introns have been eliminated from the TK-domain during the long period of the sponge's separate evolution. This view cannot be excluded considering the fact that — based on populational genetic studies - the sponge genome is particularly dynamic (Solé-Cava and Thorpe 1991). However, it is also - perhaps more - likely that introns, found in the TKdomain of existing PTKs genes, were introduced after splitting off the sponge line from the common, primitive ancestral metazoan organisms. Recent investigations from the early supporters of "genes in pieces" and the "intron early" view speak against their own theory (Stolzfus 1994). The most logical way to explain the results obtained from the analysis of the ancient G. cydonium RTK gene is to accept the "introns late" hypothesis (Orgel and Crick 1980; Cavalier-Smith 1991). Introns in the TK-domains of the TK proteins were introduced gradually during the (recent) evolution of these enzymes in the kingdom of Metazoa.

Acknowledgments. This work was supported by grants from the Deutsche Forschungsgemeinschaft (Mü 348/12-1) and the Stiftung Volkswagenwerk and the International Human Frontier Science Program (RG-333/96-M).

# References

- Agnès F, Shamoon B, Dina C, Rosnet O, Birnbaum D, Galibert F (1994) Genomic structure of the downstream part of the human FLT3 gene: exon/intron conservation among genes encoding receptor tyrosine kinase (RTK) of subclass III. Gene 145:283–288
- Arndt W (1930) Schwämme (Porifera, Spongien). In: Oppenheimer C, Pincussen L (eds) Tabulae Biologicae. W. Junk-Verlag, Berlin, pp 772–797
- Ausubel FM, Brent R, Kingston RE, Moore DD, Smith JA, Seidman JG, Struhl K (1995) Current protocols in molecular biology. John Wiley, New York
- Birchmeier C, O'Neill K, Riggs M, Wigler M (1990) Characterization

of *ROS*1 cDNA from a human glioblastoma cell line. Proc Natl Acad Sci 87:4799–4803

- Bowtell DD, Simon MA, Rubin GM (1988) Nucleotide sequence and structure of the sevenless gene of *Drosophila melanogaster*. Genes Dev 2:620–634
- Brenner S (1987) Phosphotransferase sequence homology. Nature 329:1
- Cavalier-Smith T (1991) Intron phylogeny: a new hypothesis. Trends Genet 7:145–148
- Darnell JE (1978) Implications of RNA-RNA splicing in evolution of eukaryotic cells. Science 202:1257–1260
- De Laubenfels MW (1955) Archaeocyata and porifera. In: Moore RC (ed) Treatise on invertebrate paleontology. Geol Soc Am Univ Kansas Press, Part E, pp 22–122
- Doolittle WF (1978) Genes in pieces: were they ever together? Nature 272:581–582
- Geer P, Hunter T, Lindberg RA (1994) Receptor protein-tryosine kinases and their signal transduction pathways. Annu Rev Cell Biol 10:251–337
- Gilbert W (1987) The exon theory of genes. Cold Spring Harb Symp Quant Biol 52:901–905
- Gilbert W, Marchionni M, McKnight G (1986) On the antiquity of introns. Cell 46:151–154
- Hanks SK, Quinn AM (1991) Protein kinase database: identification of conserved features of primary structure and classification of family members. Methods Enzymol 200:38–62
- Hanks SK, Quinn AM, Hunter T (1988) The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science 241:42–52
- Hao QL, Heisterkamp N, Groffen J (1989) Isolation and sequence of a novel human tyrosine kinase gene. Mol Cell Biol 9:1587–1593
- Hardie G, Hanks S (1995) The protein kinase factsbook: proteintyrosine kinases. Academic Press, London
- Henkemeyer MJ, Bennett RL, Gertler FB, Hoffmann FM (1988) DNA sequence, structure, and tyrosine kinase activity of the *Drosophila melanogaster* Abelson proto-oncogene homolog. Mol Cell Biol 8: 843–853
- Hunter T (1987) A thousand and one protein kinases. Cell 50:823-829
- Hunter T, Lindberg RA, Middlemas DS, Tracy S, Geer P (1992) Receptor protein kinases and phosphatases. Cold Spring Harb Symp Quant Biol 58:25–41
- Johnson JD, Edman JC, Rutter WJ (1993) A receptor tyrosine kinase found in breast carcinoma cells has an extracellular discoidin I-like domain. Proc Natl Acad Sci USA 90:5677–5681
- Klein R, Parada LF, Coulier F, Barbacid M (1989) trkB, a novel tyrosine protein kinase receptor expressed during mouse neural development. EMBO J 8:3701–3709
- Kozak M (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44:283–292
- Kruse M, Mikoc A, Cetkovic H, Gamulin V, Rinkevich B, Müller IM, Müller WEG (1994) Molecular evidence for the presence of a developmental gene in the lowest animals: identification of a homeobox-like gene in the marine sponge *Geodia cydonium*. Mech Ageing Dev 77:43–54.
- Lamballe F, Klein R, Barbacid M (1991) *trkC*, a new member of the trk family of tyrosine protein kinases, is a receptor for neurotrophin-3. Cell 66:967–979
- Martin-Zanca D, Hughes SH, Barbacid M (1986) A human oncogene formed by the fusion of truncated tropomyosin and protein tyrosine kinase sequences. Nature 319:43–748
- Müller WEG (1995) Molecular phylogeny of metazoa [animals]: monophyletic origin. Naturwissenschaften 82:36–38
- Orgel LE, Crick FH (1980) Selfish DNA: the ultimate parasite. Nature 284:604–607
- Orlov YA (1971) Fundamentals of palaeontology, vol II. Israel Program for the Scientific Translations, Jerusalem, p 11

- Perez JL, Shen X, Finkernagel S, Sciorra L, Jenkins NA, Gilbert DJ, Copeland NG (1994) Identification and chromosomal mapping of a receptor tyrosine kinase with a putative phospholipid binding sequence in its ectodomain. Oncogene 9:211–219
- Pfeifer K, Haasemann M, Gamulin V, Bretting H, Fahrenholz F, Müller WEG (1993a) S-type lectins occur also in invertebrates: unusual subunit composition and high conservation of the carbohydrate recognition domain in the lectin genes from the marine sponge *Geodia cydonium*. Glycobiol 3:179–184
- Pfeifer K, Frank W, Schröder HC, Gamulin V, Rinkevich B, Müller IM, Müller WEG (1993b) cDNA cloning of the polyubiquitin gene from the marine sponge *Geodia cydonium* which is preferentially expressed during reaggregation of cells. J Cell Science 106:545– 554
- Roebroek AJM, Schalken JA, Verbeek JS, van den Ouweland AM, Onnekink C, Bloemers HP, van den Ven WJ (1985) The structure of the human c-fes/fps proto-oncogene. EMBO J 4:2897–2903
- Rousset D, Agnès F, Lachaume P, André C, Galibert F (1995) Molecular evolution of the genes encoding receptor tyrosine kinase with immunoglobulinlike domains. J Mol Evol 41:421–429
- Russo WM, Lukas TJ, Cohen S, Staros VJ (1985) Identification of residues in the nucleotide binding site of the epidermal growth factor receptor-kinase. J Biol Chem 260:5205–5208
- Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chainterminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467
- Sato TN, Quin Y, Kozak CA, Audus KL (1993) Tie-1 and tie-2 define another class of putative receptor tyrosine kinase genes expressed in early embryonic vascular system. [Published erratum appears in (1993) Proc Natl Acad Sci USA 90:12056] Proc Natl Acad Sci USA 90:9355–9358
- Schäcke H, Müller WEG, Gamulin V, Rinkevich B (1994a) The Ig superfamily includes members from the lowest invertebrates to the highest vertebrates. Immunol Today 15:497–498
- Schäcke H, Rinkevich B, Gamulin V, Müller IM, Müller WEG (1994b) Immunoglobulin-like domain is present in the extracellular part of the receptor tyrosine kinase from the marine sponge *Geodia cydonium*. J Mol Recognit 7:272–276

Schäcke H, Schröder HC, Gamulin V, Rinkevich B, Müller IM, Müller

WEG (1994c) Molecular cloning of a receptor tyrosine kinase from the marine sponge *Geodia cydonium:* a new member of the receptor tyrosine kinase class II family in invertebrates. Mol Membr Biol 11:101–107

- Seino S, Seino M, Nishi S, Bell GI (1989) Structure of the human insulin receptor gene and characterization of its promoter. Proc Natl Acad Sci USA 86:114–118
- Shier P, Watt VM (1989) Primary structure of a putative receptor for a ligand of the insulin family. J Biol Chem 264:14605–14608
- Shore P, Sharrocks AD (1995) The MADS-box family of transcription factors. Eur J Biochem 229:1–13
- Shtivelman E, Lifshitz B, Gale RP, Roe BA, Canaani E (1986) Alternative splicing of RNAs transcribed from the human *abl* gene and from the *bcr-abl* fused gene. Cell 47:277–284
- Solé-Cava AM, Thorpe JP (1991) High levels of genetic variation in natural populations of marine lower invertebates. Biol J Linn Soc 44:65–80
- Stephens RM, Schneider TD (1992) Features of spliceosome evolution and function inferred from analysis of the information at the human splice sites. J Mol Biol 228:1124–1136
- Stoddard BL, Biemann HP, Koshland DE (1992) Receptors and transmembrane signaling. Cold Spring Harb Symp Quant Biol 54:1–15
- Stolzfus A (1994) Origin of introns-early or late? Nature 369:526-527
- Ullrich A, Schlessinger J (1990) Signal transduction by receptors with tyrosine kinase activity. Cell 61:203–212
- Ullrich A, Bell RJ, Chen EY, Herrera R, Petruzelli LM, Dull TJ, Gray A, Coussens L, Liao YC, Tsubokawa M (1985) Human insulin receptor and its relationship to the tyrosine kinase family of oncogenes. Nature 313:56–761
- Ullrich A, Gray A, Tam AW, Yang-Feng T, Tsubokawa M, Collins C, Henzel W, Le-Bon T, Kathuria S, Chen E (1986) Insulin-like growth factor I receptor primary structure: comparison with insulin receptor suggests structural determinants that define functional specificity. EMBO J 5:2503–2512
- Wickens M, Stephenson P (1984) Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent messenger RNA 3' end formation. Science 226:1045–1051
- Zarkower D, Stephenson P, Sheets M, Wickens M (1986) The AAUAAA sequence is required both for cleavage and for polyadenylation of Simian Virus 40 pre-mRNA in vitro. Mol Cell Biol 6:2317–2323