

Biotechnology Advances

Submitted 010910; ~7000 words

Research review article

Atypical (RIO) protein kinases from *Haemonchus contortus* - Promise as new targets for nematocidal drugs

Bronwyn E. Campbell ^a, Peter R. Boag ^{b1}, Andreas Hofmann ^{c1}, Cinzia Cantacessi ^a,
Conan K. Wang ^c, Paul Taylor ^d, Min Hu ^a, Zia-ud-Din Sindhu ^a, Alex Loukas ^e, Paul
W. Sternberg ^f, Robin B. Gasser ^{a,*}

^a *Department of Veterinary Science, The University of Melbourne, Werribee, Victoria, Australia*

^b *Department of Biochemistry and Molecular Biology, Faculty of Medicine, Nursing and Health Sciences, Monash University, Victoria, Australia*

^c *Structural Chemistry Program, Eskitis Institute for Cell & Molecular Therapies, Griffith University, Brisbane, Queensland, Australia*

^d *Institute for Structural Biology, School of Biological Sciences, The University of Edinburgh, Scotland, UK*

^e *Queensland Tropical Health Alliance, James Cook University, Cairns, Queensland, Australia*

^f *Biology Division, California Institute of Technology, Pasadena, CA, USA*

¹ Equal contributions

* Corresponding author. Tel.: +61 3 97312000; fax: +61 3 97312366.
E-mail address: robinbg@unimelb.edu.au (R.B. Gasser).

ABSTRACT

Almost nothing is known about atypical kinases in multicellular organisms, including parasites. Supported by information and data available for the free-living nematode, *Caenorhabditis elegans*, and other eukaryotes, the present article describes three RIO kinase genes, *riok-1*, *riok-2* and *riok-3*, from *Haemonchus contortus*, one of the most important parasitic nematodes of small ruminants. Analyses of these genes and their products predict that they each play critical roles in developmental pathways of parasitic nematodes. The findings of this review indicate prospects for functional studies of these genes in *C. elegans* (as a surrogate) and opportunities for the design of a novel class of nematode-specific inhibitors of RIO kinases. The latter aspect is of paramount importance, given the serious problems linked to anthelmintic resistance in parasitic nematode populations of livestock.

© 2010 Elsevier B.V. All rights reserved.

ARTICLE INFO

Keywords:

Parasite

Haemonchus contortus

RIO kinases

Structure

Relationships

Inferred function

Drug targets

1. Introduction
 2. Methodologies
 - 2.1. *Nucleic acids*
 - 2.2. *Full-length cDNAs encoding RIOKs and transcriptional profiling*
 - 2.3. *Full-length riok genes*
 - 2.4. *Bioinformatic methods*
 3. Three *riok* genes and their inferred gene products for *Haemonchus contortus*
 4. Inference of gene function and interactions based on information available for *C. elegans* and other eukaryotic organisms
 5. Three-dimensional structural modelling
 6. Application of the RIOK-1 model for the prediction of drugs *in silico*
 7. Conclusions
- Acknowledgements
- References

1. Introduction

Protein kinases are a group of enzymes that are crucial for the regulation of a wide range of cellular processes, including cell-cycle progression, transcription, DNA replication and metabolic functions. These enzymes catalyse the transfer of phosphates to serine, threonine and tyrosine residues, and thus play functional roles in reversible protein phosphorylation (Hanks et al., 1988). Based on their structure, protein kinases can be classified into eukaryotic protein kinases (ePKs) and atypical protein kinases (aPKs) (Manning et al., 2002). The aPKs have kinase activity and share limited sequence similarity to the majority of eukaryotic protein kinases. Of the 518 kinases known to be encoded in the human genome, for instance, 40 are aPKs. These molecules have been classified into 13 families or groups, one of which represents the RIO kinases (designated here as RIOKs) (Manning et al., 2002). These kinases are considered essential for life, but almost nothing is known about them for multicellular organisms (= metazoans), including nematodes (LaRonde-LeBlanc and Wlodawer, 2005a,b).

Although the primary structures (= amino acid sequences) of RIOKs are divergent from those of other protein kinase families, their conformation is similar to those of known canonical protein kinases. In yeast (*Saccharomyces cerevisiae*), RIOK-1 has been shown to have protein kinase activity *in vitro*, which is dependent upon amino acid residues recognised as being essential for kinase function (Angermayr et al., 2002). Other research has demonstrated an important role for RIOK-1 in cell-cycle progression (G1 to S transition), the regulation of the onset of anaphase and mitotic chromosome stability (Angermayr et al., 2002) as well as the processing of 20S precursor ribosomal RNA (rRNA) to the 18S species (Vanrobays et al., 2001). RIOK-2 of *S. cerevisiae* is also required for 20S rRNA processing. However, in contrast to RIOK-1, RIOK-2 appears to be localised predominantly to the nucleus (Vanrobays et al., 2001). Although RIOK-1 and RIOK-2 are known to be associated with the 20S precursor rRNA in yeast, the biological activities of both kinases do not overlap (Vanrobays et al., 2001; Geerlings et al., 2003). Recently, RIOK-2 has also been identified as an essential, late-acting 40S ribosome synthesis factor (Granneman et al., 2010). Interestingly, an RNA interference (RNAi) screen of human kinase genes identified that the knock-down of *riok-1* and *riok-2* decreased cell viability and accelerated epithelial cell migration, respectively, whereas no effect was detected for *riok-3* which is specific to metazoans (Simpson et al., 2008). A recent report (Kimmelmann et al., 2008) suggests that human RIOK-3 might be involved in human tumour cell motility and invasion, possibly through the modulation of the Rho family of GTPases. Taken together, this information indicates that RIOKs are involved in diverse and crucial biological processes in eukaryotes, but their precise roles remain to be elucidated in multicellular organisms.

RIOKs are encoded in the genome of the best-characterized metazoan, the free-living nematode *Caenorhabditis elegans* (see Manning, 2005). RNAi, which decreases messenger RNA (mRNA) levels of the targeted *C. elegans* gene, has been shown to affect predominantly embryonic and larval growth and/or development (Fraser et al., 2000; Ashrafi et al., 2003; Simmer et al., 2003; Rual et al., 2004; Sonnichsen et al., 2005). In spite of the functional importance of this molecule, there is no published information on aPKs for any related, parasitic nematodes, with the exception of *Trichostrongylus vitrinus* (order Strongylida) (see Hu et al., 2008). In the present article, we elucidate the full-length complementary cDNAs and genes of three RIOKs from *Haemonchus contortus* (an economically important blood-feeding strongylid nematode of small ruminants) and compare them with related molecules encoded in other organisms, infer the three-dimensional structures of RIOKs by comparison with known crystal structures of homologues, and assess the potential of these kinases as novel drug targets in parasitic helminths.

2. Methodologies

2.1. Nucleic acids

Genomic DNA was extracted from 50 mg of pooled *H. contortus* using a small-scale sodium dodecyl-sulphate (SDS)/proteinase K extraction procedure, followed by purification over a mini-column (Wizard Clean-Up, Promega) (Gasser et al., 2006). The specific identity and mono-specificity of the parasite material was verified by PCR-coupled, automated sequencing of the second internal transcribed spacer (ITS-2) of nuclear ribosomal DNA from genomic DNA (see Bott et al., 2009). Total RNA was extracted separately from different developmental stages (eggs, L1s, L2s, L3s and L4s [i.e., first- to fourth-stage larvae, respectively]) and sexes (females and males) of *H. contortus* (as described by Nikolaou et al., 2002). RNA yields were estimated spectrophotometrically (NanoDrop 1000), and the integrity of RNA was confirmed by detecting discrete 18S and 28S rRNA bands on ethidium bromide-stained agarose gels. Each RNA sample (~10 µg) was treated with 2 U of *DNase I* (Promega) and incubated at 37 °C for 30 min prior to heat denaturation of this enzyme (75 °C for 5 min). Single-stranded (ss) cDNA was synthesized from *DNase I*-treated total RNA (500 ng) from each developmental stage and each sex of *H. contortus* by oligo d(T)-priming using SuperScript III reverse transcriptase, as recommended by the manufacturer (Invitrogen). Nucleic acids were stored at -70 °C.

2.2. Full-length cDNAs encoding RIOKs and transcriptional profiling

For each *riok* gene, using primers designed to a number of conserved coding regions (see Supplementary Table 1), partially overlapping cDNA fragments were produced separately from total RNA from adult worms of *H. contortus* using 5'- and 3'-rapid amplification of cDNA ends (RACE) (SMART™ RACE cDNA Amplification Kit, BD Biosciences). The cDNAs were ligated separately into the pGEM-T-Easy® vector (Promega); *Escherichia coli* (strain JM109) (10⁸ colony forming units/µg) was transformed with recombinant plasmids *via* heat shock and then grown overnight at 37 °C on Luria Bertani (LB) plates containing 10 mg/ml ampicillin, 0.5 mM isopropyl-β-d-thiogalactopyranoside (IPTG) and 80 µg/ml 5-bromo-4-chloro-3-indolyl-β-galactosidase (X-gal). Plasmid DNA was isolated from recombinant clones and column-purified (Wizard®, Promega) from overnight cultures, and inserts sequenced in both directions using vector oligonucleotide primers (T7 and SP6), employing Big Dye Terminator v.3.1 chemistry in an automated ABI-PRISM sequencer (Applied Biosystems). Based on the resultant sequences, selected oligonucleotide primers (to the 5'- and 3'-termini; Supplementary Table 1) were designed to amplify the full-length cDNAs from *H. contortus*, which was subsequently cloned and sequenced by primer walking.

For the profiling of transcription, ss cDNA (~200 ng) was subjected to PCR (50 µl volume) using gene-specific primers (listed in Supplementary Table 1) under the following cycling conditions: initial denaturation at 95 °C/2 min, followed by 30 cycles of 95 °C/30 s, 55 °C/30 s and 72 °C/1 min, with a final extension of 72 °C/5 min. An additional primer set (Supplementary Table 1) was used in PCR to amplify a short region (248 bp) from the elongation factor 1 α gene (GenBank accession no. CB016314.1) as a reference control. Following PCR, an aliquot of each amplicon (10%) was examined by agarose gel electrophoresis. The specificity and identity of individual amplicons were verified by sequencing using the same primers (individually) as employed for their amplification. Samples without template (no-DNA controls) were included in each PCR run.

2.3. Full-length *riok* genes

Each of the three *riok* genes was amplified by long-PCR (BD Advantage 2, Clontech) from ~100 ng of total genomic DNA purified from pooled adult *H. contortus*, employing primers (Supplementary Table 1) located at the 5'- and 3'-termini of the cDNAs. The cycling conditions in a 2400 thermal cycler (Applied Biosystems) were: 92 °C/2 min (initial denaturation); then 92 °C/10 s (denaturation); 60 °C/30 s (annealing); 68 °C/10 min (extension) for 10 cycles, followed by 92 °C/10 s; 60 °C/30 s; 68 °C/10 min for 20 cycles, with an elongation of 10 s for each cycle, and a final extension at 68 °C/7 min. Each PCR yielded a single band upon agarose gel electrophoretic analysis. Abundant amplicons were excised from the gel, purified over a mini-spin column (Wizard® PCR-Preps, Promega), cloned into the vector pGEM®-T-Easy and then used as a template for automated sequencing, employing (separately) vector primers T7 and SP6. The sequences obtained were assembled using the CAP program (Huang et al., 1992) at the Resources Centre INFOBIOGEN (http://www.infobiogen.fr/services/analyseq/cgi-bin/cap_in.pl). The genomic sequence of each *riok* gene from *H. contortus* was compared with those of selected (closely related) *C. elegans* genes available in WormBase (release WS215; <http://www.wormbase.org/>). The exon/intron boundaries of the full-length and related full-length genes were inferred based on the alignment of the respective cDNA and genomic DNA sequences, following the GT-AG rule (Breathnach and Chambon, 1981). Where required and to ensure clarity, the prefixes *Hc-*, *Ce-* and *Af-* were used for transcripts, cDNAs, genes and gene products representing *H. contortus*, *C. elegans* and *Archaeoglobus fulgidus*, respectively.

2.4. Bioinformatic methods

Nucleotide sequences were assembled using the program CAP3 (<http://bio.ifom-firc.it/ASSEMBLY/assemble.html>). The full-length cDNA sequences were conceptually translated (six different frames) into amino acid sequences using the Baylor College of Medicine (BCM) Search Launcher (available at <http://searchlauncher.bcm.tmc.edu/seq-util/Options/sixframe.html>) and aligned using the program ClustalW (Thompson et al., 1994; available at <http://www.ebi.ac.uk/clustalw/index.html>). Sequences were compared with those available in public, non-redundant databases, using BLASTn and BLASTx algorithms (Altschul et al., 1997) available *via* the National Center for Biotechnology Information (NCBI; <http://www.ncbi.nlm.nih.gov/BLAST>), the Sanger Centre (www.sanger.ac.uk/Projects/Celegans/) and the Parasite Genome database (www.ebi.ac.uk/parasites/parasite_blast_server.html), to verify the identity of the molecules isolated. Detailed comparative analyses with *C. elegans* data in WormBase were also carried out. Given the presence of alternative transcripts of *riok* genes in *C. elegans*, the *H. contortus* genes were examined for the presence of potential alternatively spliced transcripts employing the gene prediction program AUGUSTUS (Stanke and Morgenstern, 2005; <http://augustus.gobics.de/>), utilizing data for *Brugia malayi* (available *via* NCBI) as the training set.

Protein motifs of RIOKs from *C. elegans* were classified according to their InterPro and Gene Ontology (GO; Ashburner et al., 2000) using the program InterProScan (<http://www.ebi.ac.uk/InterProScan/>; Hunter et al., 2009). The approach by Cantacessi et al. (2010) was employed to evaluate the potential of *H. contortus* and *C. elegans* RIOKs as drug targets (= 'druggability'). In brief, the InterPro domains inferred from individual RIOK proteins were compared with known domains that bind small molecules which follow the 'Lipinski rule of 5' regarding bioavailability (Lipinski et al., 1997; Hopkins and Groom, 2002). Similarly, GO terms were mapped to Enzyme Commission (EC) numbers, and a list of enzyme-targeting drugs was compiled based on

information available in the BRENDA database (www.brenda-enzymes.info; Robertson, 2005; Chang et al., 2009). The presence of known homologues linked to gene perturbation phenotypes in *Drosophila melanogaster* (using information available from FlyBase at <http://flybase.org/>; Tweedie et al., 2009) and *M. musculus* (Mouse Genome Informatics at <http://www.informatics.jax.org>; Bult et al., 2008), in addition to those in *S. cerevisiae* (Saccharomyces Genome Database at <http://www.yeastgenome.org>; Cherry et al., 1997), was also established based on sequence data available in the OrthoMCL database (www.OrthoMCL.org).

Phylogenetic analysis was conducted using Bayesian Inference (BI) (Ronquist and Huelsenbeck, 2003), employing the program MrBayes 3.1.2 (<http://mrbayes.csit.fsu.edu/index.php>). RIOK sequences from *Aspergillus* spp. (GenBank accession nos. XP_755742, XP_001208765, XP_002379813), *Cryptococcus neoformans* (XP_569448.1) and *Ixodes scapularis* (XP_002408198.1) were used as outgroups (Supplementary Tables 2-4). Posterior probabilities (pp) were calculated using 200,000 generations (ngen = 200,000; burnin = 20), employing four simultaneous tree-building chains (nchains = 4), saving every 100th tree (samplefreq = 100).

In addition to phylogenetic analysis, genetic interactions of *C. elegans* orthologues *riok-1*, *riok-2* and *riok-3* were predicted using probabilistic functional gene networking in GeneOrienteer (<http://www.geneorienteer.org/>; Zhong and Sternberg, 2006), using the recommended, stringent cut-off value of 4.6. The predicted networks were saved in a graphic display file (gdf) format and manually drawn using Adobe Illustrator CS2 (Adobe Inc). GeneOrienteer includes ~22% of the genes in *C. elegans* with known orthologues in *D. melanogaster* (vinegar fly), *S. cerevisiae* (yeast) and *Homo sapiens* (humans), and contains key information on cellular localisation, GO and expression. The computational networks make predictions regarding essentiality for diverse cellular and developmental processes. In addition, a functional classification of the interacting genes, according to their Pfam and GO terms (Ashburner et al., 2000), was conducted using the HMMER software available at <http://hmmer.janelia.org/>; Eddy, 2009).

For structural analysis and the mapping of protein 'hot spots', three-dimensional models of *Hc*-RIOK-1, *Hc*-RIOK-2 and *Hc*-RIOK-3 were generated using a comparative approach. A structure-based sequence alignment of *Hc*-RIOK amino acid sequences with those of *Af*-RIOK-1 and *Af*-RIOK-2 (from *A. fulgidus*) was generated using amino acid sequence alignments, generated manually and guided by secondary structures predicted using PSIPRED (Bryson et al., 2005). Each domain was subjected to separate fold recognition using the PSIPRED server (Bryson et al., 2005). For *Hc*-RIOK-1 and *Hc*-RIOK-3, models of the RIOK domain were generated based on the structure-based sequence alignment with *Af*-RIOK-1 (PDB accession code 1ztf), corresponding to *Hc*-RIOK-1 (113-336) and *Hc*-RIOK-3 (226-449). The model for *Hc*-RIOK-2 comprised the N-terminal winged helix domain and the RIO kinase domain (4-292), and was computed using *Af*-RIOK-2 (PDBV accession code 1zar) as a template. Twenty independent models were calculated using MODELLER (Sali and Blundell, 1993), and that with the lowest energy value was selected. The geometry was scrutinized using PROCHECK (Laskowski et al., 1993). The models were inspected visually using program O (Jones et al., 1991).

The program LIDAEUS (Wu et al., 2003) was used to search the SPECS chemical database (<http://www.specs.net>), representing 240,000 small molecular compounds with lead properties, for potential ligands of *Hc*-RIOK-1. The compounds from this database were converted to three-dimensional structures using CONCORD (TRIPOS associates) to produce force field ligand atom types suitable for van der Waals energy calculations. The site-points defining the available space in the inferred active site in the *Hc*-RIOK-1 model were calculated using a grid spacing of 0.5 Å. LIDAEUS exhaustively fits each ligand to the binding region defined. Initial orientations were obtained by fitting

ligand atoms of appropriate type on to the grid defining the binding region. Finally, a 'least-squares energy' minimization of each ligand orientation was carried out to provide a ranking score. The scoring function accounted for van der Waals, hydrophobic and H bonding interactions. An analysis of the screening results was undertaken by visual inspection, using the program SIRIUS (Buzko, 2008; <http://sirius.sdsc.edu>), as well as in-house developed software (Hofmann and Wlodawer, 2002).

3. Three *riok* genes and their inferred gene products for *Haemonchus contortus*

The full-length cDNAs (designated *Hc-riok-1*, *Hc-riok-2*, *Hc-riok-3*) were 1842, 1590 and 1434 nucleotides (nt) in length, respectively (GenBank accession nos. XXXXX-XXXXX; Table 1). The transcripts representing *Hc-riok-1*, *Hc-riok-2* and *Hc-riok-3* were reproducibly detected by reverse transcription PCR in all developmental stages examined, except for the exsheathed L3 (Fig. 1). Transcription was usually greatest in eggs, L4 female and adult male. Interestingly, transcription in female L4s was higher than in male L4s; this latter transcription pattern was essentially reversed for the female and male adults (Fig. 1).

The protein *Hc*-RIOK-1 inferred from the cDNA was 488 amino acids (aa) in length and contained the signature LVHADLSEYNTL (PS01245), characteristic of the RIOK-1 family (cf. Angermayr and Bandlow, 2002). Comparisons with sequences in non-redundant databases conducted by BLASTx analysis showed that *Hc*-RIOK-1 has significant sequence similarity/identity to related molecules from a range of organisms, including other nematodes, arthropods, vertebrates and plants, although the majority of these molecules remain to be characterized. The highest amino acid similarity was found to the protein inferred from the *riok-1* gene of *T. vitrinus* (GenBank accession no. FM209038; E-value: 0; 87% identity) and from the *riok-1* gene of *C. elegans* (sequence code M01B12.5; E-value: $1e^{-139}$; 63% identity) (see www.wormbase.org). Pairwise comparisons of amino acid sequence differences between *Hc*-RIOK-1 and selected full-length sequences representing other phyla (Supplementary Table 2) revealed sequence identities of 21-54%. The proteins *Hc*-RIOK-2 and *Hc*-RIOK-3 were 529 and 504 aa in length, respectively. Sequence comparisons of these two proteins with those of other organisms indicated most aa similarity with *C. elegans* RIOK-2 (GenBank accession no. Y105E8B.3; E-value: $3e^{-168}$; 59% identity; Supplementary Table 3) and RIOK-3 (accession no. ZK632.3; E-value: $3e^{-146}$; 55% identity; Supplementary Table 4), respectively. The protein *Hc*-RIOK-2 had 35-54% identity with full-length sequences from a range of organisms, such as other nematodes, arthropods, vertebrates and plants (see Supplementary Table 3). An alignment of this protein with RIOK-2 from a wide range of other eukaryotic organisms revealed conservation for two motifs (C-terminal catalytic domain [GxGKES] and winged helix domain) across all taxa examined. Comparison of the inferred protein *Hc*-RIOK-3 with homologues from other organisms, including other nematodes, revealed identities of 31-53% (Supplementary Table 4). Interestingly, an alignment of the latter *H. contortus* protein with other nematodes and other eukaryotes, revealed a nematode-specific aa change (ATGKES) in one of the conserved motifs (STGKES).

Phylogenetic analyses of inferred protein sequence data (Supplementary Tables 2-4) by Bayesian inference allowed the relationships of individual *H. contortus* RIOKs to selected full-length homologues from a range of other organisms, including other nematodes, to be studied (Supplementary Figs. 1-3); *Hc*-RIOK-1 clustered together with homologues from three species of *Caenorhabditis* spp. and *T. vitrinus* (CAR64255), to the exclusion of RIOK-1s from other organisms (pp = 1.00) (Supplementary Fig. 1). Similarly, *Hc*-RIOK-2 grouped together with homologues from three *Caenorhabditis* spp.

and the parasitic nematode, *B. malayi* (GenBank accession no. XP_001895435) (Supplementary Fig. 2). Although fewer full-length molecules of RIOK-3 are presently available in current databases, homologues from a range of different taxa could be compared (Supplementary Fig. 3). Based on the analysis, RIOK-3s from nematodes were distinctly different from and grouped to the exclusion of molecules from other organisms, such as mammals, fish and amphibians (pp = 1.00). These phylogenetic analyses indicated that, although there are some conserved elements in each of the three RIOKs from a range of different organisms, RIOKs from nematodes consistently grouped (with high nodal support) to the exclusion of those of other groups of eukaryotic organisms.

The structures of the three *riok* genes from *H. contortus* (Fig. 2) were inferred through separate alignments of genomic DNA sequences with cDNA sequences. The gene *Hc-riok-1* was 5697 nt in length, and comprised 16 exons and 15 introns; this gene encoded a cDNA of 1842 nt and a predicted protein of 614 aa. In contrast, the *C. elegans* gene *riok-1* was 4924 nt in size, with 7 exons and 6 introns, the cDNA (MO1B12.5a; 1521 nt) encoding a protein of 506 aa. Interrogation of WormBase indicated that two additional transcripts (representing codes MO1B12.5b.1 and MO1B12.5b.2) of *C. elegans riok-1* existed, for which the cDNAs (both 795 nt) encoded proteins of 264 aa in length. Prediction of the open reading frames (ORFs) for the *riok-1* gene of *H. contortus* indicated the existence of a potential additional transcript (1671 nt; 557 aa), although this alternative transcript has not been demonstrated experimentally. The gene in *H. contortus*, like the homologue in *C. elegans* (Y105E8B.3), was inferred to have an ORF. It was 6872 nt in length (14 exons and 13 introns) encoding a transcript of 1590 nt and a predicted protein of 529 aa. The *C. elegans* homologue (code Y105E8B.3) had two possible ORFs (4955 nt and 5253 nt, respectively). The two alternatively spliced transcripts of the *C. elegans* gene *riok-2* (2558 nt and 2739 nt, respectively) both encoded proteins of 510 aa in length. The gene *riok-3* (2910 nt) of *H. contortus* was the shortest of all three *riok* genes in this species and encoded a transcript of 1515 nt (protein 504 aa), predicted to be 1434 nt (478 aa) using AUGUSTUS. *C. elegans riok-3* had fewer exons (5) and introns (4) than *H. contortus riok-3* (with 14 exons and 13 introns).

4. Inference of gene function and interactions based on information available for *C. elegans* and other eukaryotic organisms

In *C. elegans*, *riok-1* is involved in biological processes essential for nematode viability and fertility as well as endocytosis and fat storage (Fraser et al., 2000; Ashrafi et al., 2003; Simmer et al., 2003; Rual et al., 2004; Sonnichsen et al., 2005; Balklava et al., 2007; Ceron et al., 2007; cf. www.wormbase.org; Table 2). In contrast, knowledge of the functions of the *C. elegans riok-2* and *riok-3* is limited to the observation that gene perturbation by RNAi results in lethality and developmental defects of F1 progeny, respectively (Gonczy et al., 2000; Sonnichsen et al., 2005; Table 2). Functional classification by InterPro terms of the *C. elegans* RIOKs allowed the identification of four conserved protein motifs, namely the 'RIO kinase' (InterPro code: IPR000687), 'protein kinase-like' (IPR011009), 'RIO-like' (IPR018934) and 'RIO kinase, conserved site' (IPR018935) domains (Table 2). An additional protein domain (i.e., 'RIO-2 kinase, winged helix, N-terminal'; IPR015285) was identified as being unique to RIOK-2 (Table 2). Previously, a winged helix (WTH) domain had been characterised in the N-terminal region of the yeast RIOK-2 and shown to be absent from the RIOK-1 subfamily (LaRonde-LeBlanc et al., 2005). Although this domain is structurally similar to the 'transcriptional regulator MarR and SlyA' domain (IPR000835), whose DNA binding function has been implicated in a variety of biological processes in *Escherichia coli* and *Salmonella typhimurium* (see Alekshun and Levy, 1999), GO classifications assigned the

C. elegans RIOKs to three different 'molecular function' terms, including 'ATP binding' (GO:0005524), 'catalytic activity' (GO:0003824) and 'protein serine/threonine kinase activity' (GO:0004674) (Table 2).

The *C. elegans* *rioks* were predicted to interact with a total of 246 other genes (score cut-off: 4.6; Fig. 3). Functional classification by Pfam terms and GO of the peptides encoded by the interacting genes revealed that most of them were kinases (i.e., protein kinases [PF00069], tyrosine kinases [PF07714] and lipopolysaccharide kinases [PF06293]; 40%) and phosphotransferases (PF01636; 7%). Other interactors included ribosomal proteins (PF00900; 2%) and ubiquitins (PF00179; 1.5%). As GO provides a hierarchy that unifies the descriptions of biological and molecular functions and cellular localisation (Ashburner et al., 2000), this approach was employed to classify the peptides encoded by the interacting genes. The predominant GO terms were 'protein amino acid phosphorylation' (GO:0006468; 62%) and 'intracellular signalling pathway' (GO:0023034; 5.2%) for 'biological process', 'ATP binding' (GO:0005524; 33.8%) and 'protein kinase activity' (GO:0004672; 20%) for 'molecular function' and 'intracellular' (GO:0005622; 44.5%) and 'ribosome' (GO:0005840; 23.8%) for 'cellular component' (not shown). The GO term 'protein serine/threonine kinase activity' (GO:0004648), assigned to all *C. elegans* RIOKs, could be linked to the 'non-specific' (i.e., atypical) serine/threonine protein kinase' group of molecules listed in the BRENDA database (EC:2.7.11.1; not shown). A total of 55 inhibitors were listed that specifically target this group of molecules, including the cyclin-dependent kinase inhibitors alsterpaullone and indirubin-3' monoxime (Bain et al., 2003) and the natural flavonoid kaempferol (Cho et al., 2007) (see Supplementary Table 5). Future studies could focus on the prediction and optimization of the chemical structure of the *C. elegans* RIOKs that, together with the three-dimensional modelling predicted herein for *Hc*-RIOK-1 (see below), could lead to the definition of a target protein structure to assist in the design of drugs for subsequent *in vitro* and *in vivo* testing (cf. Gasteiger, 2006; Krasky et al., 2007).

5. Three-dimensional structural modelling

In this section, all residue numbers refer to the sequence of *Hc*-RIOK-1 (see Fig. 4). Topologically, the three RIOKs differ in the N- and C-terminal domains that flank the central kinase domain. RIOK-1 usually possesses an N-terminal domain of ~100 aa residues, with few predicted secondary structure elements. *Af*-RIOK-1 from *A. fulgidus*, the only RIOK-1 for which an experimental three-dimensional structure is available (PDB accession code 1ztf) (LaRonde-LeBlanc et al., 2005b), seems to be rather unique in that it possesses a short N-terminal domain of 19 aa residues. In contrast, the extended N-terminal domain of RIOK-3 comprises >200 residues, and the secondary structure prediction indicates the presence of several folded elements that are scattered throughout this region. Thus far, no significant homology of the N-terminal domains of either RIOK-1 or RIOK-3 with known structures could be found. This finding is different from RIOK-2, in which the N-terminal domain comprises ~60 residues that are predicted to adopt ordered secondary structure elements. Indeed, the only experimental structure for any RIOK-2, determined for *Af*-RIO2 from *A. fulgidus* (PDB accession codes 1zao, 1zar) (LaRonde-LeBlanc et al., 2005a), revealed the fold of a winged helix domain, a structural feature found predominantly in nucleic acid-binding proteins (Aravind et al., 2005). The presence of this domain within RIOK-2 agrees with its reported involvement in the processing of rRNA (Vanrobays et al., 2001; Schäfer et al., 2003; Vanrobays et al., 2003).

The three RIOKs also differ in their C-terminal domains, which show no sequence identity amongst one another. For *Hc*-RIOK-1 and *Hc*-RIOK-2, these domains comprise ~150 and 240 aa residues, respectively, and some alpha-helical structure elements are predicted, embracing a highly flexible region of ~40-50 residues. *Hc*-RIOK-3 had the

shortest C-terminal domain in this respect, comprising ~50 residues which are predicted to contain two alpha-helical elements. Although no homologous proteins for the C-terminal domains of *Hc*-RIOK-1 and *Hc*-RIOK-2 could be detected, the occurrence and positioning of predicted secondary structure elements in the *Hc*-RIOK-3 C-terminal domain suggest a compact fold that might form contacts with the central kinase domain. Intriguingly, when analysing full-length *Hc*-RIOK-3, human MAP kinase kinase (PDB accession code 1s9i) and the phosphorylase kinase from *Oryctolagus cuniculus* (rabbit) (PDB accession code 1phk) were homologous proteins, with p-values of $<1e^{-4}$. Structure-based sequence alignment of *Hc*-RIOK-3 with human MAP kinase reveals that the C-terminal alpha-helical elements in *Hc*-RIOK-3 may well be accommodated in a similar manner (data not shown).

Structurally, the defining criterion for RIOKs is the presence of a shared central kinase domain, which resembles the fold of eukaryotic protein kinase (ePK) domains. The ePK fold is characterised by 12 structural features (known as subdomains, indicated with Roman numerals; see Supplementary Fig. 4) (Hanks and Hunter, 1995), adopting a bi-lobal shape as an overall structure. The smaller N-terminal lobe is comprised of subdomains I-IV and is involved in binding and orienting the phosphate donor (ATP or GTP). The larger C-terminal lobe includes subdomains VIa-XI, and fulfils the tasks of peptide substrate binding and the initiation of the transfer of the phosphate group. Compared with the ePK fold, the RIOK domain lacks three of these features, namely the activation loop (also called APE-loop or subdomain VIII, involved in peptide substrate recognition), and subdomains X and XI.

The essential features of the catalytic domain of ePKs comprise conserved regions that are responsible for binding of the phosphate donor (subdomains I-III), metal binding (DFG-loop, subdomain VII), substrate binding (subdomains V, VIII) and phosphoryl transfer (subdomain VIb). The catalytic loop (subdomain VIb) harbours two invariant residues (Asp and Asn) that are part of the RIOK signature motif. The invariant Asp273 is situated in this loop and most likely acts as the catalytic residue, which extracts the proton from the substrate hydroxyl group. Importantly, a conserved, basic residue in ePKs at position 275 (lysine or arginine) is thought to play a key role during the transfer of the γ -phosphate by neutralising its negative charge. This particular mechanism seems to be completely abolished in RIOKs, most of which possess a serine residue at this position and no other basic residue in the vicinity of the active site that could fulfil this role.

Glu276 as well as Asn278 are both strictly conserved residues, which are located in the catalytic loop. Glu276 participates in substrate binding and also forms a direct interaction with the ribose moiety of the nucleotide. Asn278 stabilises the conformation of the catalytic loop through hydrogen bonds to the backbone of Asp273 as well as binding one of the two metal ions.

The correct positioning of the nucleotide by residues in the P-loop (subdomain II) as well as the invariant residue Lys157 and the presence of three invariant side chains (Glu141, Asn278 and Asp290) enable octahedral coordination of each of the two metal ions (Mg^{2+} or Mn^{2+}). The coordination sphere of each of the cations is provided by three contacts from the phosphate donating nucleotide, two from the protein and one from a water molecule. Notably, the conserved DFG-loop of ePKs in the metal binding region is variable among RIOKs, which only shows strict conservation for Asp290. Furthermore, many ePKs possess a conserved cysteine residue preceding the DFG-loop (Hur et al., 2008). This position represents a conserved isoleucine in RIOKs (Ile289).

The most variable elements in the kinase catalytic domains are the hinge and the flexible loop between $\beta 3$ and αC . When viewing the RIOK catalytic domain from the top (see Fig. 4), these two segments with low amino acid sequence conservation and flexible structure are located at the opposite ends of the central substrate and nucleotide binding groove. Variation and low structural order in these regions can be tolerated,

since these loop regions are not involved in the packing of the three-dimensional structure of the kinase domain. In ePKs, the hinge (subdomain V) harbours a conserved residue that participates in binding the substrate peptide. Although the flexible loop of ePKs might contain a helical element (α B), there are also members of the family, such as Cdk2 and Erk2, from which this helix is absent.

For RIOKs, it can be speculated that both the hinge and flexible loop act as a two-point substrate recognition or binding feature that orients the substrate and maintains its position during the catalytic phosphoryl transfer. Intriguingly, an auto-phosphorylation site has been reported for *Af*-RIOK-2 at position 165 (Ser128 in *Af*-RIOK-2), which is located in the flexible loop (LaRonde-LeBlanc et al., 2005a). The structural and functional consequence of this post-translational modification remains elusive, but it is tempting to speculate a regulatory role, in which a conformational change within the flexible loop affects catalytic efficiency and/or substrate binding. Notably, the serine residue at position 165 is conserved for RIOK-2.

Although the phosphorylation of *Af*-RIOK-1 has been reported to be essential for catalytic activity (Angermayr and Bandlow, 2002), the auto-phosphorylation of this protein at the residue in position 187 (Ser108 in *Af*-RIOK-1), which is located at the very N-terminal end of the α C helix and represents the base of the catalytic groove, does not seem to be relevant functionally, since the Ser108Ala mutant of *Af*-RIOK-1 has shown wild-type catalytic activity (LaRonde-LeBlanc et al., 2005b). Moreover, the aa at position 187 varies among different RIOKs, even within a group, further questioning its relevance as a regulating position.

As identified in our alignment of members of all three RIO kinases (Supplementary Fig. 4), as well as in an alignment of RIOK-1 members from a variety of organisms, including the parasitic nematode *T. vitrinus* (see Hu et al., 2008), Ser165 is part of the conserved dipeptide motif TS present in the flexible loop. It is tempting to propose that this serine is indeed a conserved (auto-)phosphorylation site in RIOKs, which warrants future evaluation.

6. Application of the RIOK-1 model for the prediction of drugs *in silico*

In a first attempt to probe the active site of RIOKs, with a view toward drug discovery, we conducted an *in silico* screen using the homology model of *Hc*-RIOK-1 employing the SPECS database. The top 12 binding compounds identified from this screen are listed in Table 3. Interestingly, four of these 12 compounds possess a carbohydrate moiety. For the compounds ranked third and tenth, a second binding mode was observed and ranked in positions 8 and 11, respectively, indicating an increased likelihood that these molecules would display productive interactions in an *in vitro* assay.

The hydrogen-bond interactions between the compounds identified and the protein model involved several conserved side chains in the active site that belong to functionally important elements (including P-loop, catalytic loop, metal binding loop), such as Glu141, Lys157, Asp273, Asn278, Asp290 and Gln293. However, all top-12 compounds were also involved in interactions with residues that are not conserved and specific for *H. contortus* RIOK-1, including Tyr159, Ser275 and Ser292. Such interactions are considered of crucial importance for the design of specific inhibitors for RIOKs.

Using 12 known kinase effectors, as listed in the BRENDA database entry for 'serine/threonine protein kinase', we searched further the list of top 500 molecules from the *in silico* screen to find compounds with similar chemical structures to the known effectors by using an algorithm to identify the maximum common subgraph between two individual compounds. Two compounds in the top-500 list, namely at ranks 9 and 256, from the *in silico* screen displayed similarity to the protein kinase inhibitor emodin (emodin, Schuttgelb), which is an anthraquinone found in several plants and formerly used as a laxative or purgative (National Toxicology Program, 2001). The similarity is

based on the anthraquinone scaffold present in emodol, and the compounds ranked at 9 and 256. The fact that a similarity match occurred to a compound among the top-10 could make this scaffold a useful starting point for future drug design studies.

The compounds ranked at positions 40, 81, 241 and 256 possessed some structural similarity to the flavonoids apigenin and kaempferol, two very similar kinase effectors listed in the BRENDA database. Flavone compounds share the annealed ring system with naphtha- or anthraquinones. Apigenin is an inhibitor of cytochrome P450 2C9 (Dayong et al., 2009) and has been recognised as an anti-cancerous effector from vegetables and fruits (Ferreira et al., 2006). Kaempferol is a component of tea, and many fruits and vegetables (Park et al., 2006), and has also been recognised to possess cancer-protective effects (Cho et al., 2007; Nöthlings et al., 2007). Probably the most interesting compound identified in the virtual screen was prunitrin, ranked at position 2. Prunitrin is a naturally occurring isoflavonoid in clover (*Trifolium* species) and the Rosaceae family (*Prunus* species) (Cooke and Fletcher, 1974), and thus may be a common component of the diet of animals. Intriguingly, prunitrin combines the presence of a naphthaquinone scaffold (emodol identity score: 0.2716; see Table 3) and a carbohydrate moiety, which are both functionalities that are repeatedly represented in the top 12 results from the *in silico* screen.

7. Conclusions

This article provides new insights into three different atypical (RIO) kinases in *H. contortus*, one of the economically most important parasites of livestock. Based on bioinformatic and phylogenetic analyses, these kinases are proposed to be novel drug targets. Currently, computational approaches (e.g., Krasky et al., 2007; Caffrey et al., 2009; Doyle et al., 2010) are increasingly being used to assess the potential of key genes/gene products as novel drug targets in parasitic worms. In addition, structure-based virtual screening has been proven to be useful in the identification of compounds able to inhibit the activity of molecules whose three dimensional structure had been established using homology models (reviewed by Villoutreix et al., 2007). For instance, *in silico* docking for ~200,000 compounds (i.e., ChemDiv) into the binding site of an homology model of a human BCR-ABL tyrosine kinase (which is known to play a crucial role in the pathogenesis of chronic myeloid leukaemia; Deininger et al., 2000) led to the identification of 15 compounds selected for biological testing, eight of which were demonstrated to significantly inhibit tumor cell growth (Peng et al., 2003). In another study (Vangrevelinghe et al., 2003), novel and selective inhibitors of protein casein kinase II were identified by *in silico* docking of an homology model of CK2 with a subset of 400,000 molecules available in the Novartis database (Vangrevelinghe et al., 2003). Although numerous examples of protein-ligand interaction studies and drug design using *in silico* approaches are described in the literature (cf. Villoutreix et al., 2007; Cavasotto and Phatak, 2009; Hammami and Fliss, 2010), and computational structure prediction methods are cost- and time-effective in the absence of experimental structures, the success of these approaches depends heavily on the accuracy of the model predicted and on the sequence similarities between the protein used as a template and the homologous sequence(s) (see Cavasotto and Phatak, 2009).

Based on modelling, structural comparison of the three RIOKs shows that the RIOK domain harbouring the catalytic site is a well-conserved fold among parasitic nematodes, in particular between *H. contortus* and *T. vitrinus*. However, despite this fold, there are several aa substitutions in functionally important, conserved secondary structure elements (Supplementary Figure 4), whose impact can only be assessed from three-dimensional structures determined experimentally. Future structural studies will therefore be needed to reveal the particular binding modes of ligands, particularly the phosphate-donating nucleotides, in order to provide a solid basis for structure-based drug design. Furthermore, several mechanistic aspects of RIOKs are poorly understood, also requiring detailed structural information. Our current working model assumes that the two flexible elements in the RIOK domain, the hinge and the flexible loop (see Fig. 4), serve as docking points for the substrate and might undergo conformational change in the substrate-bound state. Such a process may be further aided by phosphorylation of Ser165 (*Hc*-RIOK-1 numbering), which is located in the flexible loop and seems to be a conserved residue for RIOKs. Clearly, crystal structures of substrate-bound and phosphorylated nematode RIOKs will assist in elucidating the molecular biology of these proteins.

Here, we attempted to probe the active site of one representative of RIOKs in *H. contortus*, *Hc*-RIOK-1, using the structural model obtained by comparison with the known structures of *Af*-RIOK-1 and *Af*-RIOK-2. Although preliminary, the *in silico* study revealed two interesting observations. First, of the top 500 compounds identified in the screen, molecules that possessed a naphthaquinone, anthraquinone or flavone scaffold occurred repeatedly; this finding is of particular importance, given that protein kinase inhibitors, such as the anthraquinone emodol and the flavonoid kaempferol, are already known. Secondly, a number of compounds identified by *in silico* screening possess a carbohydrate moiety which might be a second feature of a future drug candidate, targeting RIOKs. The natural compound prunitrin, identified as the second compound in the *in silico* screen, combines both features and thus constitutes a very promising starting point for future drug discovery. Importantly, all of the top-12 compounds were inferred to interact not only with conserved residues in the RIOK domain, but also with residues that are specific for *H. contortus* RIOK-1, thus providing clues as to how to design specific inhibitors.

Despite the relative evolutionary conservation of members of the RIOK family, there is essentially no understanding of how these proteins function. From the limited data available for *C. elegans*, yeast and human cell lines (Fraser et al., 2000; Vanrobays et al., 2001; Angermayr et al., 2002; Ashrafi et al., 2003; Simmer et al., 2003; Rual et al., 2004; Sonnichsen et al., 2005; Simpson et al., 2008), it is known that the RIOKs are essential for cellular viability. However, developmental and tissue-specific functions of RIOKs can only be revealed by studying them in an experimentally tractable, multicellular system. By taking a "whole animal" approach, it should be possible to reveal the developmental and tissue specificity of RIOK function(s) in *C. elegans*. The technical arsenal available to the researcher working on *C. elegans* permits the study of biological pathways *in vivo*, and provides a well-suited eukaryotic system. Given that the complete genome sequence of *C. elegans* has been determined, a well-defined model system is now in place to assess, in great detail, the functionality of homologous genes from related, parasitic nematodes *via* genomic and phenotypic analyses. This aspect is of paramount importance, because most parasitic nematodes cannot be propagated and maintained effectively in culture systems *in vitro*. Therefore, it should be possible to selectively explore how RIOKs are associated with parasite development in the mammalian host, and make connections between the parasite genes (its genome) and their function in *C. elegans* (e.g., by RNAi; Fire, 2007) based on physical appearance and behaviour (i.e., phenome). Bridging the gap between genome and proteome will also provide unique insights into processes that are both shared by and unique to

parasitic and free-living nematodes. This approach, coupled to the use of cutting-edge technologies, will allow us to assay the complex nematode system and gain insight into why parasitic organisms are so pervasive and cause chronic infections in their hosts, in spite of a host immune response. This link between the genome, phenome and proteome will not only enable us to obtain a comprehensive understanding of the function of RIOKs, but also of the whole organism and how it interacts with its environment. Therefore, we propose that an interdisciplinary approach should be taken to explore this interesting group of aPKs. Advances in this area could have significant implications for the development of nematode-specific RIOK inhibitors, with prospects for developing an entirely new class of nematocides. This aspect has major significance, given the substantial problems associated with anthelmintic resistance in parasitic nematodes of livestock.

Acknowledgements

Funding from the Australian Research Council (ARC) is gratefully acknowledged (RBG).

References

- Alekshun MN, Levy SB. The mar regulon: multiple resistance to antibiotics and other toxic chemicals. *Trends Microbiol* 1999;7:410-3.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389-402.
- Angermayr M, Bandlow W. Rio1, an extraordinary novel protein kinase. *FEBS Lett* 2002;524:31-6.
- Angermayr M, Roidl A, Bandlow W. Yeast Rio1p is the founding member of a novel subfamily of protein serine kinases involved in the control of cell cycle progression. *Mol Microbiol* 2002;44:309-24.
- Aravind L, Anantharaman V, Balaji S, Babu, MM, Iyer LM. The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev* 2005;29:231-62.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;25:25-9.
- Ashrafi K, Chang FY, Watts JL, Fraser AG, Kamath RS, Ahringer J, et al. Genome-wide RNAi analysis of *Caenorhabditis elegans* fat regulatory genes. *Nature* 2003;421:268-72.
- Bain J, McLauchlan H, Elliott M, Cohen P. The specificities of protein kinase inhibitors: an update. *Biochem J* 2003;371:199-204.
- Balklava Z, Pant S, Fares H, Grant BD. Genome-wide analysis identifies a general requirement for polarity proteins in endocytic traffic. *Nat Cell Biol* 2007;9:1066-73.
- Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL. The Pfam protein families database. *Nucleic Acids Res* 2000;28:263-66.
- Bott NJ, Campbell BE, Beveridge I, Chilton NB, Rees D, Hunt PW, et al. A combined microscopic-molecular method for the diagnosis of strongylid infections in sheep. *Int J Parasitol* 2009;39:1277-87.
- Breathnach R, Chambon P. Organization and expression of eucaryotic split genes coding for proteins. *Ann Rev Biochem* 1981;50:349-83.
- Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS, Jones DT. Protein structure prediction servers at University College London. *Nucleic Acids Res* 2005;33:W36-8.
- Bult CJ, Epping JT, Kadin JA, Richardson JE, Blake JA. The mouse genome database (MGD): mouse biology and model systems. *Nucleic Acids Res* 2008;36:D724-8.
- Caffrey CR, Rohwer A, Oellien F, Marhofer RJ, Braschi S, Oliveira G, et al. A comparative chemogenomics strategy to predict potential drug targets in the metazoan pathogen, *Schistosoma mansoni*. *PLoS One* 2009;4:e4413.
- Cantacessi C, Mitreva M, Jex AR, Young ND, Campbell BE, Hall RS, et al. Massively parallel sequencing and analysis of the *Necator americanus* transcriptome. *PLoS Negl Trop Dis* 2010;4:e684.
- Cavasotto CN, Phatak SS. [Homology modeling in drug discovery: current trends and applications](#). *Drug Discov Today* 2009;14:676-83.
- Ceron J, Rual JF, Chandra A, Dupuy D, Vidal M, van den Heuvel S. 2007. Large-scale RNAi screens identify novel genes that interact with the *C. elegans* retinoblastoma pathway as well as splicing-related components with synMuv B activity. *BMC Dev Biol* 2007;7:30.
- Chang A, Scheer M, Grote A, Schomburg I, Schomburg D. BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. *Nucleic Acids Res* 2009;37:D588-92.
- Cherry JM, Ball C, Weng S, Juvic G, Schmidt R, Adler C, et al. Genetic and physical maps of *Saccharomyces cerevisiae*. *Nature* 1997;387:67-73.
- Cho YY, Yao K, Kim HG, Kang BS, Zheng D, Bode AM, et al. Ribosomal S6 kinase 2 is a

- key regulator in tumor promoter induced cell transformation. *Cancer Res* 2007;67:8104-12.
- Conder GA, Johnson SS. Viability of infective larvae of *Haemonchus contortus*, *Ostertagia ostertagi*, and *Trichostrongylus colubriformis* following exsheathment by various techniques. *J Parasitol* 1998;82:100e2.
- Cooke RG, Fletcher, RAH. Isoflavonoids III - constituents of *Cotoneaster* species. *Aust J Chem* 1974;27:1377-79.
- Dayong S, Wang Y, Zhou YH, Guo Y, Wang J, Zhou H, et al. Mechanism of CYP2C9 inhibition by flavones and flavonols. *Drug Metabol Dispos* 2009;37:629-4.
- Deininger MW, Vieira S, Mendiola R, Schultheis B, Goldman JM, Melo JV. BCR-ABL Tyrosine kinase activity regulates the expression of multiple genes implicated in the pathogenesis of chronic myeloid leukemia. *Cancer Res* 2000;60:2049-55.
- DeLano WL. The PyMOL molecular graphics system 2002 (<http://www.pymol.org>).
- Doyle MA, Gasser RB, Woodcroft BJ, Hall RS, Ralph SA. Drug target prediction and prioritization: using orthology to predict essentiality in parasite genomes. *BMC Genomics* 2010;11:222.
- Eddy SR. [A new generation of homology search tools based on probabilistic inference](#). *Genome Inform* 2009;23:205-11.
- Ferreira CV, Gusto GZ, Sousa AC, Queiroz KC, Zambuzzi WF, Aoyama H, Peppelenbosch MP. Natural compounds as a source of protein phosphatase inhibitors; application to the rational design of small-molecule derivatives. *Biochimie* 2006;88:1859-73.
- Fire AZ. Gene silencing by double-stranded RNA. *Cell Death Differ* 2007;14:1998-2012.
- [Fraser AG](#), [Kamath RS](#), [Zipperlen P](#), [Martinez-Campos M](#), [Sohrmann M](#), [Ahringer JA](#). Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* 2000;408:325-30.
- Gasser RB, Hu M, Chilton NB, Campbell BE, Jex AR, Otranto D, et al. Single-strand conformation polymorphism (SSCP) for the analysis of genetic variation. *Nature Protoc* 2006;1:3121-8.
- Gasteiger J. Chemoinformatics: a new field with a long tradition. *Anal Bioanal Chem* 2006;384:57-64.
- Geerlings TH, Faber AW, Bister MD, Vos JC, Raué HA. Rio2p, an evolutionarily conserved, low abundant protein kinase essential for processing of 20S Pre-rRNA in *Saccharomyces cerevisiae*. *J Biol Chem* 2003;278:22537-45.
- [Gonczy P](#), [Cassin E](#), [Hannak E](#), [Kirkham M](#), [Pichler SC](#), [Flohrs K](#), et al. Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* 2000;408:331-6.
- Granneman S, Petfalski E, Swiatkowska A, Tollervey D. Cracking pre-40S ribosomal subunit structure by systematic analyses of RNA-protein cross-linking. *The EMBO J* 2010;29:2026-36.
- Hammami R, Fliss I. Current trends in antimicrobial agent research: chemo- and bioinformatics approaches. *Drug Discov Today* 2010;15:540-6.
- Hanks SK, Hunter T. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J* 1995;9:576-96.
- Hofmann A, Wlodawer A. PCSB - a programme collection for structural biology and biophysical chemistry. *Bioinformatics* 2002;18:209-10.
- Holliday JD, Hu CY, Willett P. Grouping of coefficients for the calculation of inter-molecular similarity and dissimilarity using 2D fragment bit-strings. *Comb Chem High Throughput Screen* 2002;5:155-66.
- Hopkins AL, Groom CR. The druggable genome. *Nat Rev Drug Discov* 2002;1:727-30.
- Hu M, LaRonde-LeBlanc N, Sternberg PW, Gasser RB. *Tv*-RIO1 - an atypical protein kinase from the parasitic nematode *Trichostrongylus vitrinus*. *Parasit Vectors* 2008;1:34.
- Huang X. A contig assembly program based on sensitive detection of fragment overlaps.

- Genomics 1992;14:18-25.
- Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, et al. [InterPro: the integrative protein signature database](#). Nucleic Acids Res 2009;37:D211-15.
- Hur W, Velentza A, Kim S, Flatauer L, Jiang X, Valente D, et al. Clinical stage EGFR inhibitors irreversibly alkylate Bmx kinase. Bioorg Med Chem Lett 2008;18:5916-19.
- Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and location of errors in these models. Acta Crystallogr A 1991;47:110-9.
- Kimmelman AC, Hezel AF, Aguirre AJ, Zheng H, Paik JH, Ying H, et al. Genomic alterations link Rho family of GTPases to the highly invasive phenotype of pancreas cancer. Proc Natl Acad Sci U S A 2008;105:19372-7.
- Krasky A, Rohwer A, Schroeder J, Selzer PM. A combined bioinformatics and cheminformatics approach for the development of new antiparasitic drugs. Genomics 2007;89:36-43.
- LaRonde-LeBlanc N, Wlodawer A. A family portrait of the RIO kinases. J Biol Chem 2005a;280:37297-300.
- LaRonde-LeBlanc N, Wlodawer A. The Rio kinases: an atypical protein kinase family required for ribosome biogenesis and cell cycle progression. Biochim Biophys Acta 2005b;1754:14-24.
- Laskowski R, MacArthur M, Moss DS, Thornton JM. PROCHECK: A program to check the stereochemical quality of protein structures. J Appl Cryst 1993;26:283-91.
- Lipinski C, Lombardo F, Dominy B, Feeney P. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 1997;23:3-25.
- Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. Science 2002;298:1912-34.
- Manning G. Genomic overview of protein kinases. WormBook 2005;1-19.
- National Toxicology Program. NTP Toxicology and Carcinogenesis Studies of EMODIN (CAS NO. 518-82-1) Feed Studies in F344/N Rats and B6C3F1 Mice. Natl Toxicol Program Tech Rep Ser 2001;493:1-278.
- Nikolaou S, Hartman D, Presidente PJ, Newton SE, Gasser RB. HcSTK, a *Caenorhabditis elegans* PAR-1 homologue from the parasitic nematode, *Haemonchus contortus*. Int J Parasitol 2002;32:749-58.
- Nöthlings U, Murphy SP, Wilkens LR, Henderson BE, Kolone LN. Flavonols and pancreatic cancer risk. Am J Epidemiol 2007;166:924-31.
- Park JS, Rho HS, Kim DH, Chang IS. Enzymatic Preparation of kaempferol from green tea seed and its antioxidant activity. J Agric Food Chem 2006;54:2951-56.
- Peng H, Huang N, Qi J, Xie P, Xu C, Wang J, et al. [Identification of novel inhibitors of BCR-ABL tyrosine kinase via virtual screening](#). Bioorg Med Chem Lett 2003;13:3693-9.
- Robertson JG. [Mechanistic basis of enzyme-targeted drugs](#). Biochemistry 2005;44:5561-71.
- Ronquist F, Huelsenbeck JP. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 2003;19:1572e4.
- Rual JF, Ceron J, Koreth J, Hao T, Nicot AS, Hirozane-Kishikawa T, et al. Toward improving *Caenorhabditis elegans* phenome mapping with an ORFeome-based RNAi library. Genome Res 2004;14:2161-8.
- Sali A, Blundell T. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234:779-815.
- Schäfer T, Strauss D, Petfalski E, Tollervey D, Hurt E. The path from nucleolar 90S to cytoplasmic 40S pre-ribosomes. EMBO J 2003;22:1370-80.
- [Simmer F, Moorman C, van der Linden AM, Kuijk E, van den Berghe PVE, Kamath FS, et al. Genome-wide RNAi of *C. elegans* using the hypersensitive *rrf-3* strain reveals novel gene functions. PLoS Biol 2003;1:77-84.](#)
- Sonnichsen B, Koski LB, Walsh A, Marschall P, Neumann B, Brehm M, et al. Full-genome

- RNAi profiling of early embryogenesis in *Caenorhabditis elegans*. *Nature* 2005;434:462-9.
- Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res* 2005;33:W465-7.
- Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673-80.
- Tweedie S, Ashburner M, Falls K, Leyland P, McQuilton P, Marygold S, et al. FlyBase Consortium. FlyBase: enhancing *Drosophila* Gene Ontology annotations. *Nucleic Acids Res* 2009;37:D555-9.
- Vangrevelinghe E, Zimmermann K, Schoepfer J, Portmann R, Fabbro D, Furet P. [Discovery of a potent and selective protein kinase CK2 inhibitor by high-throughput docking](#). *J Med Chem* 2003;46:2656-62.
- Vanrobays E, Gleizes P, Bousquet-Antonelli C, Noaillac-Depeyre J, Caizergues-Ferrer M, Gélugne J. Processing of 20S pre-rRNA to 18S ribosomal RNA in yeast requires Rrp10p, an essential non-ribosomal cytoplasmic protein. *EMBO J* 2001;20:4204-13.
- Vanrobays R, Gelugne J, Gleizes P, Caizergues-Ferrer M. Later cytoplasmic maturation of the small ribosomal subunit required RIO proteins in *Saccharomyces cerevisiae*. *Mol Cell Biol* 2003;23:2083-95.
- Villoutreix BO, Renault N, Lagorce D, Sperandio O, Montes M, Miteva MA. Free resources to assist structure-based virtual ligand screening experiments. *Curr Protein Pept Sci* 2007;8:381-411.
- Wu SY, McNae I, Kontopidis G, McClue SJ, McInnes C, Stewart KJ, et al. Discovery of a novel family of CDK inhibitors with the program LIDAEUS: structural basis for ligand-induced disordering of the activation loop. *Structure* 2003;11:399-410.
- Zhong W, Sternberg PW. Genome-wide prediction of *C. elegans* genetic interactions. *Science* 2006;311:1481-4.

Table 1. Features of the three *riok* genes from *Haemonchus contortus* (*Hc-*) and *Caenorhabditis elegans* (*Ce-*), including accession numbers, predicted alternative transcripts, lengths of genomic and cDNA sequences, numbers of introns and exons, and the predicted lengths of inferred proteins

Gene	Accession number	Length (bp)	Number of introns	Number of exons	Accession number (cDNA)	Length (bp)	Predicted amino acids
<i>Hc-riok-1</i>	<u>XXXXX</u>	5697	15	16	<u>XXXXX</u>	1842	614
<i>Hc-riok-1a</i> ^a	-	5697	14	15	-	1671	557
<i>Ce-riok-1</i>	MO1B12.5	11477	6	7	MO1B12.5	1521	506
<i>Ce-riok-1a</i>	MO1B12.5b.1	3320	6	7	MO1B12.5b.1	795	264
<i>Ce-riok-1b</i>	MO1B12.5b.1	3242	6	7	MO1B12.5b.1	795	264
<i>Hc-riok-2</i>	<u>XXXXX</u>	6872	13	14	<u>XXXXX</u>	1590	529
<i>Ce-riok-2</i>	Y105E8B.3.1	5253	2	3	Y105E8B.3.1	1590	510
<i>Ce-riok-2</i>	Y105E8B.3.2	4955	2	3	Y105E8B.3.2	1590	510
<i>Hc-riok-3</i>	<u>XXXXX</u>	2910	13	14	<u>XXXXX</u>	1515	504
<i>Hc-riok-3</i> ^a	-	2910	12	13	-	1434	477
<i>Ce-riok-3</i>	ZK632.3.1	2739	4	5	ZK632.3.1	1533	510
<i>Ce-riok-3</i>	ZK632.3.2	2558	4	5	ZK632.3.2	1533	510

^a Putative alternative transcript.

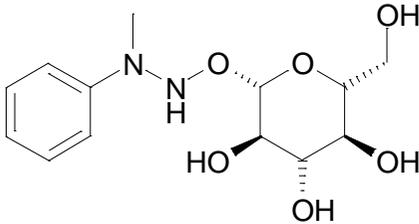
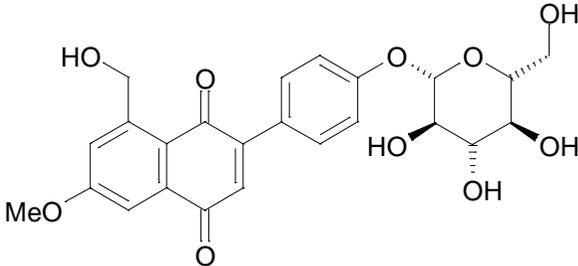
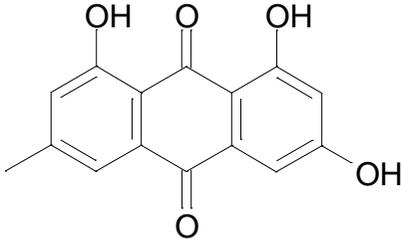
Table 2. Functional annotation of the three RIO kinases from *Haemonchus contortus*, including InterPro domain codes, gene ontology data, inferred functions and RNAi phenotypes. Helminth species for which either partial or full-length *riok*/RIOK sequences have been identified to date are also listed.

<i>C. elegans</i> gene name (gene code/ Wormbase gene ID)	Protein description	InterPro domains (InterPro codes)	Gene ontology		RNAi phenotypes ^a (alphabetical)	References	Inferred function	Other helminths ^b (alphabetical)
			Biological process	Molecular function				
<i>riok-1</i> (M01B12.5a/WBGene00019698)	Serine/threonine-protein kinase RIOK-1	RIO kinase (IPR000687) RIO kinase, conserved site (IPR018935) RIO-like kinase (IPR018934) Serine/threonine-protein kinase Rio1 (IPR017407)	Embryonic development; Growth; Genitalia development; Larval development; Receptor-mediated endocytosis; Reproduction	ATP binding; Catalytic activity; Protein serine/threonine kinase activity	Emb, embryonic defects, Gro, Let, Lva, Lvl, Ptv, reduced brood size, Sck, small, Ste	Fraser et al. (2000); Ashrafi et al. (2003); Simmer et al. (2003); Rual et al. (2004); Sonnichsen et al. (2005); Balklava et al. (2007); Ceron et al. (2007)	Entry into S phase and exit from mitosis (inferred from <i>S. cerevisiae</i>); Processing of late 18S rRNA (inferred from <i>S. cerevisiae</i>)	<i>Fhe, Hco, Mch, Ovi, Ptr, Sra, Tci, Tco, Tvi, Xin</i>
<i>riok-2</i> (Y105E8B.3/WBGene00013688)	Serine/threonine protein kinase RIOK-2	RIO kinase (IPR000687) RIO kinase, conserved site (IPR018935) RIO-like kinase (IPR018934) RIO2 kinase, winged helix, N-terminal (IPR015285) Lipopolysaccharide kinase (IPR010440)	Embryonic development; Protein amino acid phosphorylation	ATP binding; Catalytic activity; Protein serine/threonine kinase activity	Emb	Sonnichsen et al. (2005)	Processing of late 18S rRNA (inferred from <i>S. cerevisiae</i>)	<i>Csi, Fhe, Hco, Gpa, Hgl, Mch, Min, Nam, Mpa, Ptr, Tco</i>
<i>riok-3</i> (ZK632.3/WBGene00014012)	Serine/threonine protein kinase RIOK-3	RIO kinase (IPR000687) RIO kinase, conserved site (IPR018935) RIO-like kinase (IPR018934) Serine/threonine-protein kinase Rio3 (IPR017406)	Embryonic development	ATP binding; Catalytic activity; Protein serine/threonine kinase activity	Embryonic and postembryonic development variant	Gonczy et al. (2000)	Unknown	<i>Hco, Nam, Ode, Ppa, Sst, Tco, Xin</i>

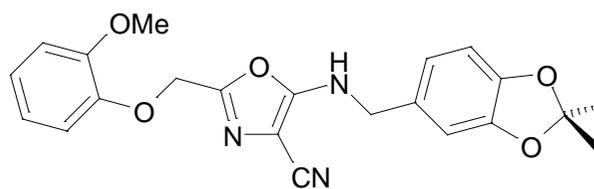
^aRNAi phenotypes: embryonic lethal (Emb), slow growth (Gro), adult lethal (Let), larval arrest (Lva), larval lethal (Lvl), protruding vulva (Pvl), sick (Sck), sterile (Ste).

^bOther helminths: *Clonorchis sinensis* (*Csi*), *Haemonchus contortus* (*Hco*), *Fasciola hepatica* (*Fhe*), *Globodera pallida* (*Gpa*), *Heterodera glycines* (*Hgl*), *Meloidogyne chitwoodi* (*Mch*), *Meloidogyne incognita* (*Min*), *Meloidogyne paranaensis* (*Mpa*), *Necator americanus* (*Nam*), *Oesophagostomum dentatum* (*Ode*), *Opisthorchis viverrini* (*Ovi*), *Parastrongyloides trichosuri* (*Ptr*), *Pristionchus pacificus* (*Ppa*), *Strongyloides ratti* (*Sra*), *Strongyloides stercoralis* (*Sst*), *Teladorsagia circumcincta* (*Tci*), *Trichostrongylus colubriformis* (*Tco*), *Trichostrongylus vitrinus* (*Tvi*), *Xiphinema index* (*Xin*).

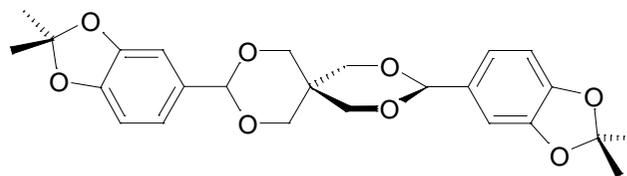
Table 3. Results from an *in silico* screen for ligands of *Hc*-R1OK-1. The top 12 compounds identified in the screen are shown, together with hits ranked at positions 40, 81, 236, 241 and 256. The score recorded accounted for hydrogen bond donor and acceptor qualities as well as van der Waals interactions, thus resembling a binding enthalpy. Residues of *Hc*-R1OK-1 engaged in hydrogen bonding interactions with the compounds are listed in the fourth column. The lead-likeness was assessed using the criteria of logP, logS, number of hydrogen bond donors, acceptors and rotatable bonds; the maximum score is thus '+++++'. The identity score between the compounds ranked and known kinase effectors was calculated as a Tanimoto coefficient (Holliday et al., 2002) by finding the largest common subgraph, with a score between 0 (nothing in common) and 1 (identical).

<i>In silico</i> screening result						Known serine/threonine kinase effector	
Rank	Identification (ID)	Score	Interactions with	Lead-like	Structure	Identity score	Structure
1	AE-562/12222311	-58.33	E141, K157, Y159, D273, D290, S292, Q293	++++			
2	AG-690/33068027 Prunitrin	-55.50	E141, K157, Y159, R169, D273, S275, Y277, N278, D290, Q293	+++		0.27	

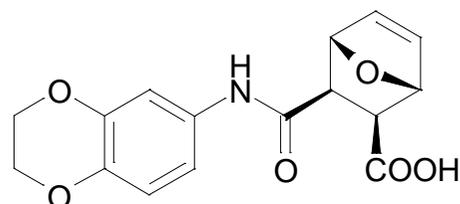
3 AF-399/41669003 -54.83 K157, D273, S275 , Y277, +++++
N278, D290, Q293



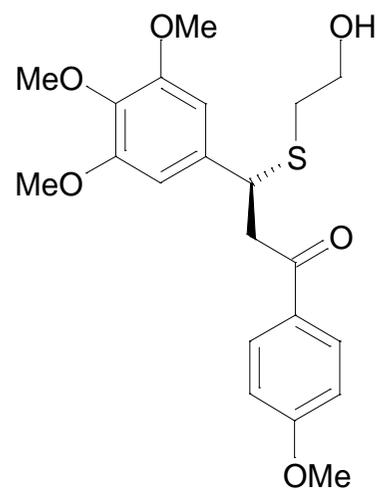
4 AG-690/11063014 -53.14 E141, Y159, R169, D273, +++++
S275, Y277, N278, D290,
S292 , Q293



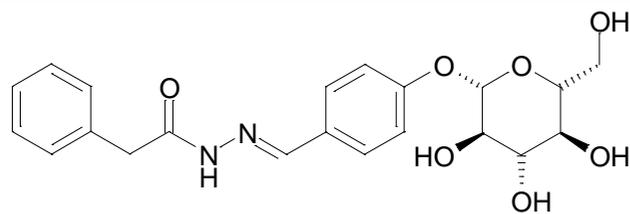
5 AK-968/11992092 -53.12 E141, K157, Y159, R169, +++++
D273, Y277, N278, D290,
Q293



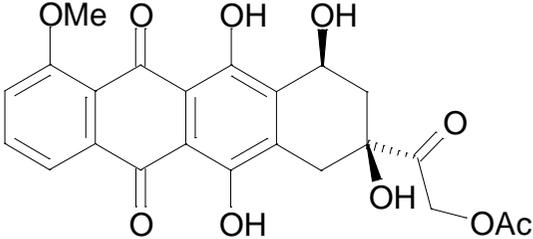
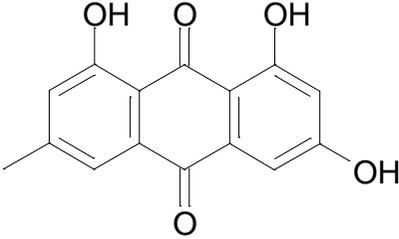
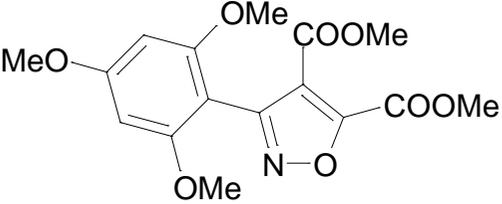
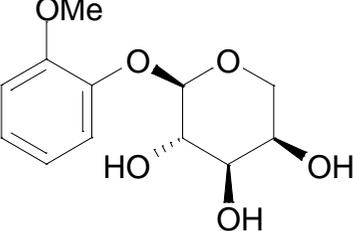
6 AF-399/40826342 -52.07 S137, E141, K157, E199, N278, D290, Q293 +++

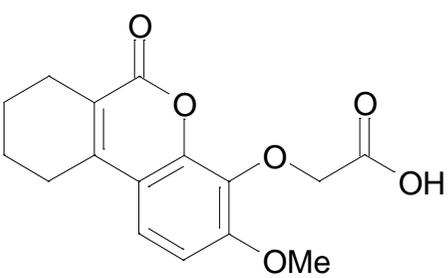
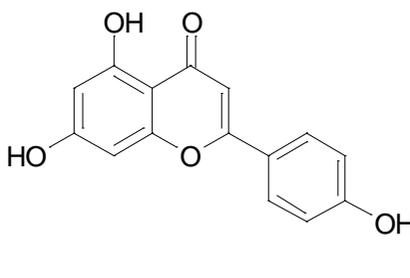
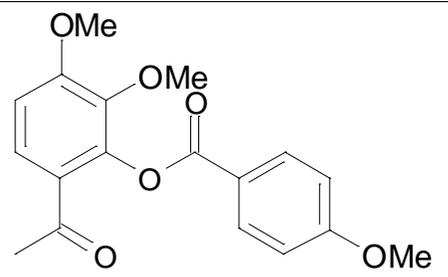
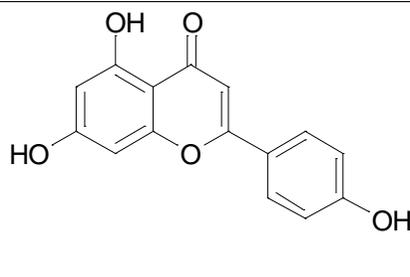
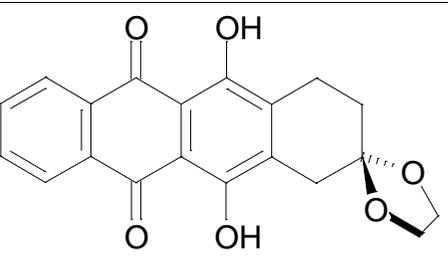
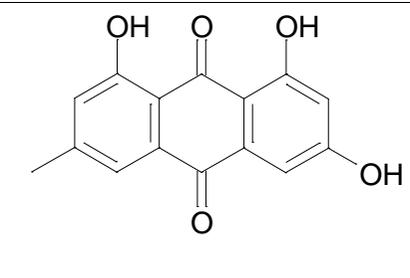


7 AG-205/14655024 -51.90 E141, K157, Y159, R169, Y277, D290, S292, Q293 ++++

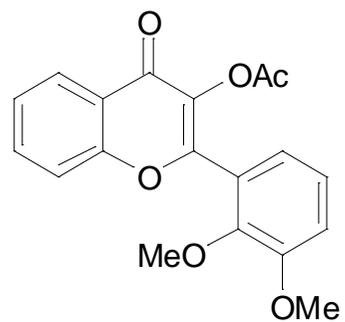


8 same compound as rank 3 -51.72 K157, D273, S275, Y277, N278, D290, Q293 ++++

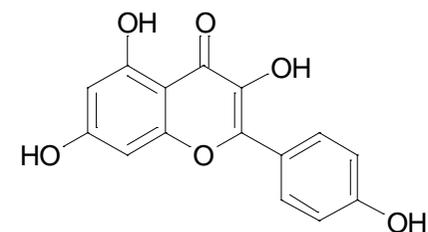
9	AG-650/41069241	-51.54	E141, K157, Y159, D273, S275, N278, D290, Q293, N312	+++		0.56	
10	AP-006/40679673	-51.37	E141, K157, Y159, H271, D273, N278, D290, S292, Q293	++++			
11	same compound as rank 10	-51.37	E141, K157, Y159, H271, D273, N278, D290, S292, Q293	++++			
12	AG-690/32520056	-59.66	E141, D273, D290, S292, Q293	++++			

40 AP-906/42853619 -46.24	E141, K157, H271, N278, D290, Q293 +++++		0.40	
81 AO-423/14105406 -44.31	E141, K157, D273, S275, Y277, N278, D290, Q293 +++++		0.76	
236 AE-508/36397055 -41.46	E141, K157, D273, D290, Q293 +++++		0.64	

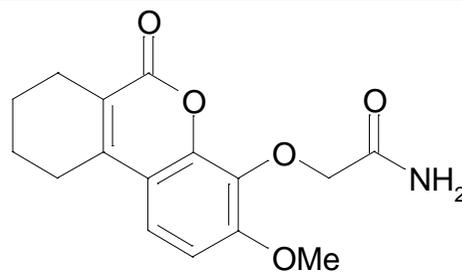
241 AA-768/30961027 -41.39 E141, K157, Y159, R169, +++++
N278, D290, S292
, Q293



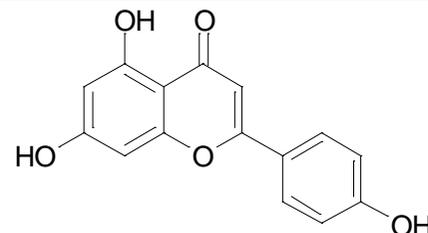
0.64



256 AP-906/42853620 -41.20 E141, R169, H271, N278, +++++
D290, Q293



0.39



LEGENDS TO FIGURES

Fig. 1. Transcriptional profiles of *riok-1*, *riok-2* and *riok-3* in different developmental stages of *Haemonchus contortus* (*Hc*-), determined by reverse transcription PCR analysis. First-, second-, third- and fourth-stage larvae (L1, L2, L3 and L4, respectively) as well as adult stage (A); female (F) and male (M). A region (248 bp) of the elongation factor 1 α gene (*elf1- α*) was used as a reference control in PCR.

Fig. 2. Diagrammatic representation of the genomic organizations of the genes *riok-1*, *riok-2* and *riok-3* from *Haemonchus contortus* (*Hc*-) and orthologues from *Caenorhabditis elegans* (*Ce*-). The organization of each gene was determined by aligning the cDNA and genomic DNA sequences, with intron-exon boundaries being defined using the GT-AG rule (Breathnach and Chambon, 1981). Black boxes represent exons, whilst lines represent introns. Numbers above the boxes indicate the size of exons (in nucleotides), whereas numbers below the lines indicate the intron sizes.

Fig. 3. Probabilistic functional gene network analysis for *Caenorhabditis elegans* genes *riok-1*, *riok-2* and *riok-3* (see <http://www.wormbase.org/>).

Fig. 4. Structure models of RIOK-1, RIOK-2, and RIOK-3 from *Haemonchus contortus* (panels A-C, respectively) inferred by comparative modelling using the structures of *Af*-RIOK-1 (PDB 1ztf) and *Af*-RIOK-2 (1zar) as simultaneous templates. Twenty independent models were generated using MODELLER (Sali and Blundell, 1993) and, for each kinase, the model with the lowest energy was selected. There are two flexible elements (shown in yellow: hinge and flexible loop) at diametral positions of the active site. Colour code for other structural elements: P-loop (green); catalytic loop (red); metal binding loop (turquoise); winged helix domain (grey). The figure was prepared using the program PyMOL (DeLano, 2002).

LEGENDS TO SUPPLEMENTARY TABLES

Supplementary Table 1. List of oligonucleotide primers used for the isolation and sequencing of the genes *riok-1*, *riok-2* and *riok-3* from *Haemonchus contortus* (*Hc*). All sequences are 5'- to 3'-; the position/s refer/s to the last nucleotide at the 3'-end of the gene (g) and/or cDNA (c).

Supplementary Table 2. RIOK-1 amino acid (aa) sequences (and relevant information from BLASTx analysis) used in the analysis of the phylogenetic relationship of RIOK-1 from *Haemonchus contortus* to related molecules.

Supplementary Table 3. RIOK-2 amino acid (aa) sequences (and relevant information from BLASTx analysis) used in the analysis of the phylogenetic relationship of RIOK-2 from *Haemonchus contortus* to related molecules.

Supplementary Table 4. RIOK-3 amino acid (aa) sequences (and relevant information from BLASTx analysis) used in the analysis of the phylogenetic relationship of RIOK-3 from *Haemonchus contortus* to related molecules.

Supplementary Table 5. Compounds predicted to inhibit the function of RIOK-1, RIOK-2 and/or RIOK-3 of *Haemonchus contortus*.

LEGENDS TO SUPPLEMENTARY FIGURES

Supplementary Fig. 1. Phylogenetic tree displaying the relationships of RIOK-1 from *Haemonchus contortus* with homologues from a range of species, including *Caenorhabditis elegans*, *C. briggsae*, *Trichostrongylus vitrinus*, *Brugia malayi* (nematodes); *Arabidopsis thaliana*, *Nicotiana tabacum*, *Oryza sativa* (plants); *Homo sapiens*, *Macaca fascicularis*, *Pan troglodytes*, *Canis familiaris*, *Rattus norvegicus*, *Mus musculus* (other mammals); *Danio rerio* (zebrafish) *Drosophila melanogaster* (vinegar fly) and *Dictyostelium discoideum* (slime mould). Accession numbers identify individual sequences representing individual species (see Table 2). *Aspergillus* spp. (fungus) (GenBank accession nos. XP_755742, XP_001208765 and XP_002379813) represent the outgroups. Numbers above the branches are the pp-values inferred from the Bayesian inference method used.

Supplementary Fig. 2. Phylogenetic tree displaying the relationships of RIOK-2 from *Haemonchus contortus* with homologues from a range of species, including *Caenorhabditis elegans*, *C. briggsae*, *Trichostrongylus vitrinus*, *Brugia malayi* (nematodes); *Arabidopsis thaliana*, *Nicotiana tabacum*, *Oryza sativa* (plants); *Homo sapiens*, *Macaca fascicularis*, *Pan troglodytes*, *Canis familiaris*, *Rattus norvegicus*, *Mus musculus* (other mammals); *Danio rerio* (zebrafish) *Drosophila melanogaster* (vinegar fly) and *Dictyostelium discoideum* (slime mould). Accession numbers identify individual sequences representing individual species (see Table 3). *Cryptococcus neoformans* (yeast) (GenBank accession no. XP_569448.1) represents the outgroup. Numbers above the branches are the pp-values inferred from the Bayesian inference method used.

Supplementary Fig. 3. Phylogenetic tree displaying the relationships of RIOK-1 from *Haemonchus contortus* with proetin homologues from a range of species, including *Caenorhabditis elegans*, *C. briggsae*, *Trichostrongylus vitrinus*, *Brugia malayi* (nematodes); *Arabidopsis thaliana*, *Nicotiana tabacum*, *Oryza sativa* (plants); *Homo sapiens*, *Macaca fascicularis*, *Pan troglodytes*, *Canis familiaris*, *Rattus norvegicus*, *Mus musculus* (other mammals); *Danio rerio* (zebrafish) *Drosophila melanogaster* (vinegar fly) and *Dictyostelium discoideum* (slime mould). Accession numbers identify individual sequences representing individual species (see Table 4), whilst *Ixodes scapularis* (tick) (GenBank accession no. XP_002408198.1) was used as the outgroup. Numbers above the branches are the pp-values inferred from the Bayesian inference method used.

Supplementary Fig. 4. Structure-based sequence alignment of the RIO kinases from *Haemonchus contortus* (*Hc*-) with RIOK-1 and RIOK-2 from *Archaeoglobus fulgidus* (*Af*-), for which three-dimensional structures have been determined. The secondary structure elements of *Hc*-RIOKs were determined using PSIPRED (Bryson et al., 2005); alpha-helices are indicated in green and beta-strands in red. The location of structural elements as known from ePKs as well as the "subdomain" assignment is shown underneath each alignment block. Numbers in the first line of each alignment block show the amino acid residue position of RIOK-1 from *H. contortus*. Abbreviations for individual residues highlights are * - catalytic residue, M = metal-binding, A = nucleotide-binding, M = hydrophobic pocket, P = auto-phosphorylation site (at residues shown in bold).