Chapter 4

Display Cloning

4.1 Overview

In the previous chapter the derivatisation of three natural products is described to establish them as probes for the bioaffinity isolation of their cellular targets.
In this chapter, we first present our results from control experiments utilising biotinylated FK506 to validate the affinity selection method. The FK506-FKBP (FK506 Binding Protein) interaction is particularly strong and therefore comprises a perfect tool to validate our libraries and the selectivity of our method[361,362]. FKBP is a small cytosolic protein that has previously been isolated using affinity chromatography[362] and phage display[63,91]. Consequently, the remaining three probes (biotinylated manzamine, daptomycin and artesunate) were screened against various human cDNA libraries and one bacterial gDNA library to isolate cellular targets and off-targets. The results from the affinity selection experiments are discussed in this chapter.

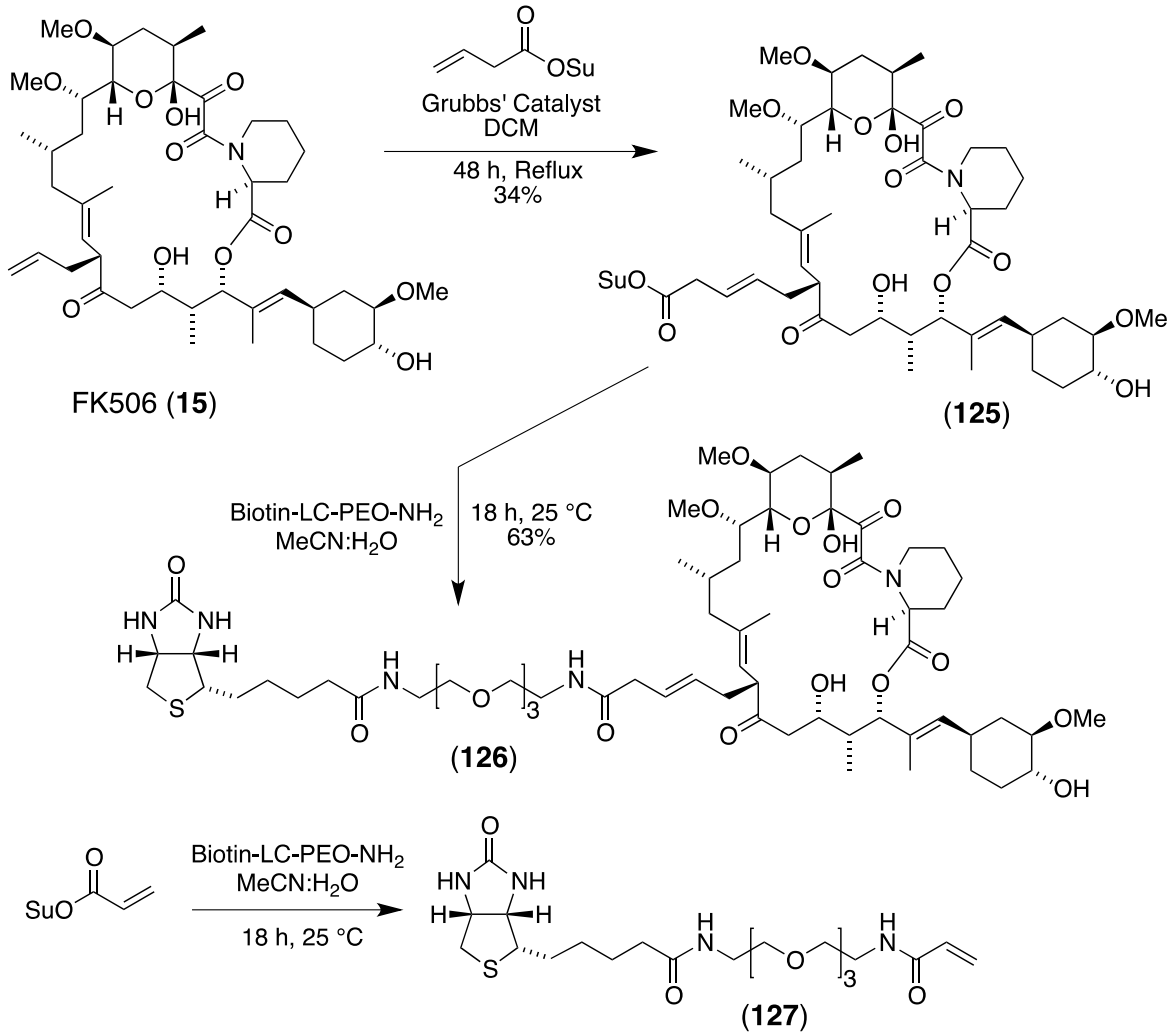
4.2 FK506

4.2.1 Introduction

FK506 (tacrolimus) is an immunosuppressant drug isolated from the fermentation of *Streptomyces tsukubaensis*[363]. It is a macrolide alkaloid and its main use follows organ transplantation in order to reduce the activity of the patient's immune system to allay the risk of organ rejection. It reduces peptidyl-prolyl isomerase (PPIase) activity by binding to the immunophilin FK506 binding protein 1a (FKBP1a), thereby creating a new complex. This FKBP1a-FK506 complex interacts with and inhibits calcineurin thus inhibiting both T-lymphocyte signal transduction and IL-2 transcription[364]. FKBPs are found almost universally across species, both in prokaryotes and eukaryotes and together with the cyclophyllins constitute the immunophilin protein family. FKBPs are involved in several biochemical processes including protein folding, receptor signalling, protein trafficking and transcription. Their primary function is assumed to be the catalysis of cis-trans isomerisation of proline amide bonds, which has been proposed as a rate-limiting step in protein folding[364,365]. FKBP1a (identical to FKBP12) is a 12 kDa cytosolic protein, which is the major cellular target for FK506 with a K_d ~0.4 nM. In addition to its high affinity for tacrolimus, FKBPs have been shown to retain their FK506 binding ability, even when displayed on the surface of T7 phage particles[63,65]. For these reasons, biotinylated FK506 was used as a positive control, to confirm our ability to isolate cellular targets for biotinylated natural products.

4.2.2 Results and discussion

The FK506 probe (**126**) has previously been synthesised in our lab (Scheme 10)[91] and was available along with the control probe (**127**), which contains only the allyl side chain of FK506.



Scheme 10: Synthesis of biotin-FK506 (126) affinity probe and biotin-acrylate (127) control[91]

To generate an affinity support for display cloning, control probe (**127**) and FK506-biotin (**126**) were immobilised on neutravidin-coated polystyrene (PS) microtitre plates. A T7 phage-displayed human brain cDNA library was then subjected to seven rounds of selection, first applying the phage library to the control probe to clear non-specific binding and then transferring the cleared library to FK506-biotin wells to selectively capture FKBP expressing phages. The library was cloned using the T7Select10-1 vector, so phage particles will display a maximum of one encoded protein on the surface, most phages express no protein so should be lost in the washing steps.

The control probe was utilised to remove phages displaying proteins capable of binding to avidin, biotin, the TEG linker or the plastic plate. Therefore, fresh phage lysate was pre-incubated in the control plate for 1 h and consequently transferred to the FK506-derivatised plate for 2 h incubation at room temperature. While a longer incubation time would increase the probability of the target phages diffusing to the

immobilised molecules, it would also increase the risk of the displayed proteins being degraded by enzymes in the cell lysate.

If longer incubations (e.g. overnight) are required, it is advisable to perform these at 4 °C in the presence of protease inhibitors, to minimise protein degradation. Following incubation, the phage solutions were aspirated and the wells were washed with phage wash buffer (PWB). For the first three rounds of selection, low stringency washes (< 10 s) were applied to prevent loss of moderate affinity binders that are present in low copy numbers. The proportion of phages displaying the target protein(s) presumably increases with each successive round of selection and the stringency of the washing step can be increased accordingly (maximum of 5 × 12 sec). Finally, phages retained by the affinity support were released non-specifically by treatment with a denaturing solution of SDS. The eluate constituted the next sublibrary, which was diluted tenfold with 2xYT and stored at 4 °C overnight. The extended cooling resulted in precipitation of a portion of the SDS. This is useful as, although T7 phages are robust and can withstand harsh conditions, high concentrations of SDS can inhibit bacterial cell growth in subsequent rounds of selection. The following morning, a small aliquot of the sublibrary was diluted 1:1000 with log phase *E. coli* (BLT5615; over-expressing the T7 coat protein gp10A) and then incubated at 37 °C until lysis was observed (typically 1-2 h). This procedure was repeated until seven rounds of selection had been completed. Typically 5-15 rounds of selection are required because the first few washes are of such low stringency that little purification occurs. As stated, the aim was simply not to loose low abundance clones.

Quantifying the overall number of phage particles eluted from the plate of each round of selection indicates the degree of specific enrichment that has occurred. To measure the phage titre, firstly a small volume of the SDS eluate from each selection round was serially diluted from 10^{-1} to 10^{-6}. An aliquot of each dilution was then dropped onto an agar plate containing a thin layer of *E. coli*-infected top agarose. After a short incubation period clear plaques became visible against the opaque bacterial lawn. The dilutions containing a countable number of plaques (5-50) were used to determine the phage particle titre of each selection round (see Figure 39) and this data used to determine if the selection had worked.

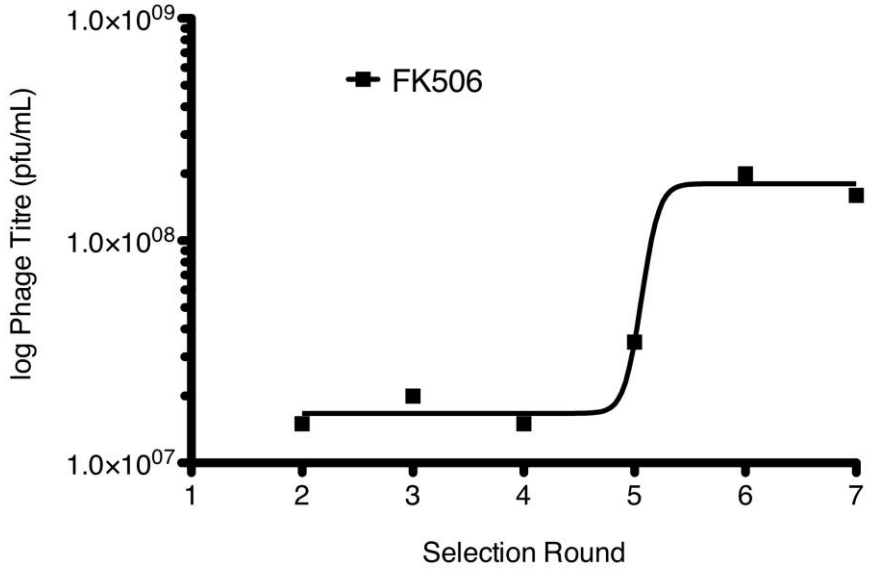


Figure 39: Phage titre of each round of selection using human brain cDNA phage library and biotinylated FK506 immobilised on neutravidin-coated plates fitted to a sigmoidal curve (Dynafit).

The human brain library showed a 10-fold increase in phage titre after seven rounds of selection. This suggests that the biotinylated FK506 affinity probe was successful in selectively enriching one or more clones from the cDNA library by round 6. Due to the lack of stringency, no increase in titre was observed till round 5 and the selection was complete by round 7. The slight drop of titre in round 7 is likely due to an extensive washing step in the final round.

PCR amplification was performed to analyse the DNA inserts from the phages surviving each selection round. Two sets of primers were used, one consisting of generic T7 primers, which amplify DNA inserts containing any gene, the other one of a mixture of 4 specific FKBP primers (FKBP1a, FKBP1b, FKBP2 and FKBP3). All amplified DNA was then subjected to gel electrophoresis (see Figure 40).

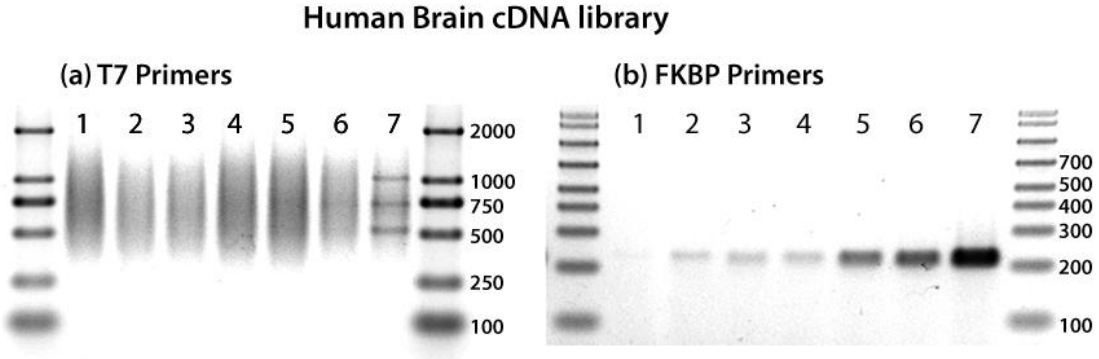


Figure 40: Agarose gel electrophoresis of phage DNA inserts amplified by PCR from human brain cDNA library after seven rounds of selection with biotin-FK506 immobilised on a neutravidin-coated polystyrene plate. (a) 1.5% agarose gel and generic T7 primers. (b) 3% SFR agarose gel and specific primers for FKBP1a, FKBP1b, FKBP2 and FKBP3 (mixed).

Over the first five rounds of selection, the DNA amplified using the generic T7 primers appeared as a smear on the gel from 350 bp to 1500 bp, indicating the presence of a wide range of different sized DNA inserts. Starting in round 6 and more obviously in round 7, three dark bands between 500 bp and 1100 bp became apparent, indicating the enrichment of DNA inserts coding for specific binding proteins. When the PCR of the DNA inserts was performed with the specific FKBP primers, already by the second round of selection a faint band appeared at about 230 bp, which gained in intensity over the consecutive 5 rounds of selection. This increase of intensity, combined with the results from the generic T7 primers based PCR, provided strong evidence that we successfully enriched at least one FKBP clone as a binding partner for the biotin-FK506 (**126**) affinity probe. Merely from its size, it was impossible to determine which FKBP was represented by the band around 230 bp, as three of the four different FKPB primers all give similarly sized PCR products (FKPB1a = 216 bp, FKBP1b = 231 bp, FKBP2 = 234 bp). PCR of the derived sublibraries with each individual specific FKBP primer would be useful, but not necessary for the aim of this study.

Additionally, to determine the ratio between FKBP and non-FKBP containing clones, single plaques of the last round of selection were isolated and analysed by PCR as follows: the fresh phage lysate of round 7 was serially diluted and an aliquot of a 10^{-7} dilution was used to infect an *E. coli*-containing molten top agarose. The molten solution containing both bacteria and phage were poured onto standard agar plates, left to set for 20 min at room temperature, and incubated at 37 °C until clear plaques were visible against the opaque lawn of bacteria. Eight random plaques were picked and their DNA inserts amplified using the generic T7 primers.

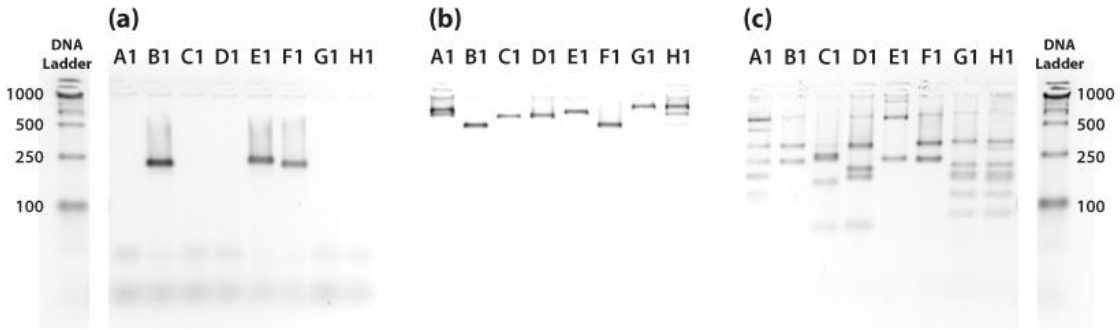


Figure 41: Agarose gel electrophoresis (3% SFR agarose) of PCR products obtained from single plaques after seven rounds of selection with biotinylated FK506 immobilised on a neutravidin-coated polystyrene plate. DNA inserts were (a) amplified with FKBP primer mix or (b) T7 generic primers; (c) products from *(a)* were enzymatically digested with the restriction endonuclease *Hin*fI.

Agarose gel electrophoresis was performed to separate the PCR products (see Figure 41a), to reveal that three out of the eight clones contained FKBP1 or FKBP2 genes. Two of those carried the similar DNA inserts, whereas the third clone (E1) differed in size. To further investigate the similarity of the isolated clones, the PCR product of each clone was enzymatically digested with a frequent base cutter (*Hin*fI) to generate a unique fingerprint of each insert. The digested PCR products were separated by gel electrophoresis (see Figure 41c). FKBP containing clones B1 and F1 gave identical fingerprints, as did G1 and H1. B1 and E1 clones were amplified (generic T7 primers) and sequenced (see Figure 42) to show these were FKBP1a and FKBP1b respectively. Phage clone B1 encoded for a protein sequence incorporating the entire human FKBP1a protein with a 21-amino acids reader sequence. The DNA insert of clone E1 is in excellent homology with the FKBP1b gene, but the sequencing results were only reliable for approximately the first 270 bp but considering that the overall length of the insert is at least 500 bp (see Figure 40a), we expected the entire gene (901 bp) plus some 3'-UTR to be present.

The sequencing results of clone G1 revealed a homology with tetra-trico-peptide repeat domain 25 (TTC25). Tetra-trico-peptide repeat (TPR) domains are found in numerous proteins, where they serve as interaction modules and multiprotein complex mediators. TPRs can be found in all kingdoms of life and regulate diverse biological processes, such as organelle targeting and protein import, vesicle fusion, and biomineralisation[366]. The cDNA insert covers the last 130 bp of the approximately 2 kb long coding sequence and further 180 bp of 3' UTR. The phage insert also contains a 700 bp long cDNA sequence, which does not result in any homology in the BLAST search. The phage-displayed protein encoded out of frame (-2) with the cDNA insert of TTC25, and was limited to an 8 aa short sequence (SADRNQKN). This suggests this clone to be a non-specific binder. It is surprising to see this clone enriched throughout the biopanning process. A possible explanation for its presence is an insufficient stringency of the washing steps. Harsher washing conditions may result in removal of this clone from the affinity support.

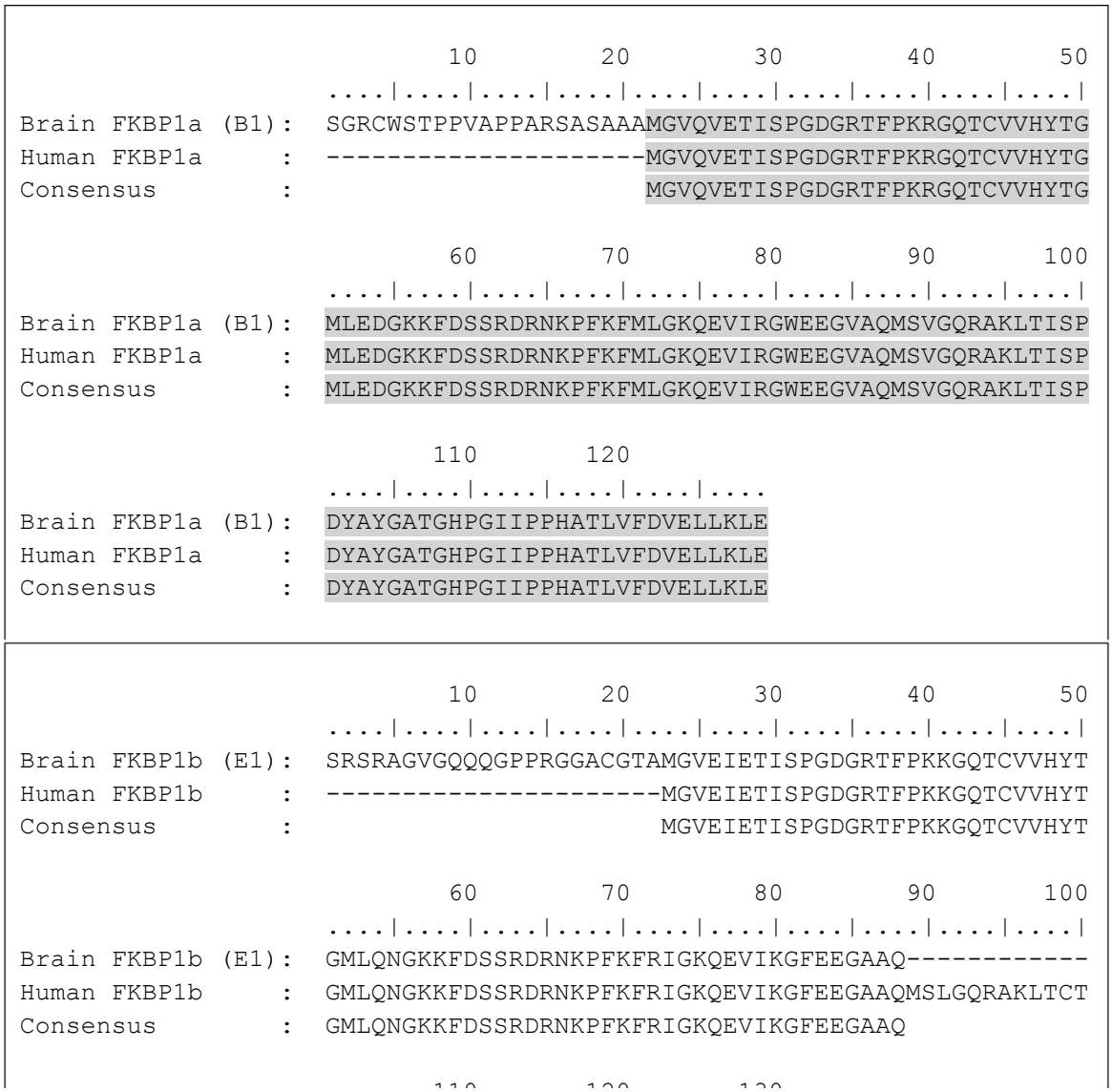


Figure 42: Alignment of T7 phage-displayed FKBP1a and FKBP1b protein sequences with the human analogues (UniProtKB: P62942 and P68106)

However, if only one single phage particle of this clone was to remain on the solid support and eluted with the specific binders, this particular clone would have an advantage during subsequent replication compared to he FKBP coding clones. For clone G1, the coding of the alien protein ends after encoding only 8 aa. In comparison, the FKBP1a-displaying clones encode for an alien sequence of 129 aa and thus may be outnumbered by the clones not caught up in expression of alien proteins.

The presence of background binders is a clear shortcoming experienced in many chemical proteomics methods (see Section 1.1.2.2). This requires a good balance in the means to remove non-specific binders. However, for this study, the isolation of two different FKBP clones from the myriad present in the library (10^6-10^7) provides an important proof of concept for the applied method. We have thereby shown that

we are able to isolate cellular receptors from a T7 phage-displayed cDNA library, using a small molecule immobilised on the neutravidin-coated surface of a 96-well microtitre plate. It should thus be possible to identify protein-binding partners for other small molecules.

4.3 Manzamine A

4.3.1 Introduction

Manzamine A is a member of the marine polycyclic manzamine alkaloids and carries the highest potential of this class of alkaloids for therapeutic relevance[367]. Its bioactivity ranges from antiprotozoal, antimicrobial (antimalarial), cytotoxic (anticancer) and is effective against neuroinflammation (see Section 3.2.3.1.1). The biochemical pathways and cellular targets for this highly interesting compound are not yet fully understood. The described inhibition of glycogen synthase kinase 3β (GSK-3β) is currently a major lead. However, the *in vitro* GSK-3β inhibition (IC_{50} = 10.2 μM) compares poorly to other GSK-3β inhibitors and cannot explain the compound's wide range of potent bioactivity. Thus, further cellular targets are assumed, but have not yet been described.

Reverse chemical proteomics is the ideal tool to help elucidate multiple targets and off-targets for this complex compound, as it allows for rapid discovery of the cognate drug-receptor pairs. To allow for isolation of the most avid protein binding partners, we utilise a wide selection of cDNA libraries derived from both normal and diseased cells of various human tissues, representing the major focus in this experiment to isolate human cellular targets of manzamine A. Given the extensively described antimicrobial effects of manzamine, we included the one gDNA library of a Gram-negative bacterium available to us.

4.3.2 Results and discussion

8-hydroxymanzamine (**92**) was biotinylated with a long, hydrophilic TEG linker to give the biotin-manzamine probe (**98,** see Scheme 4), which was immobilised on a neutravidin-coated polystyrene (PS) microtitre plate as described in the previous sections. 2-hydroxycarbazole was derivatised with the same linker to give biotin-carbazole (**99**, see Scheme 4), acting as a control probe by mimicking the carboline moiety of the manzamine alkaloids, and was immobilised on a second neutravidin-coated PS microtitre plate.

Six T7 phage-displayed human cDNA libraries (normal colon; Alzheimer's brain; cancerous cells of breast, colon, liver and lung tissue) and one gDNA library of the Gram-negative *Pseudomonas stutzeri* were then subjected to nine rounds of selection using the manzamine carrying affinity support for biopanning. All seven libraries

were cloned using the T7Select10-3 vector and therefore will display 5-15 copies of the encoded protein on the surface of each phage particle. The selections were carried out simultaneously to Section 0. In the first three rounds of selection, a very brief wash was applied to avoid loss of weak binders of low copy number. After round three, the stringency of the washes was gradually increased with each successive round, leading up to the particular rigorous washes used in rounds eight and nine to help remove more non-specific binders. To analyse sublibraries of phages surviving each selection round, we applied generic T7 primers for PCR amplification and separated the resulting PCR products by gel electrophoresis (see Figure 43 and Figure 44). No convergence was observed in any of the human libraries but one gene began to dominate in the bacterial library (see Figure 44).

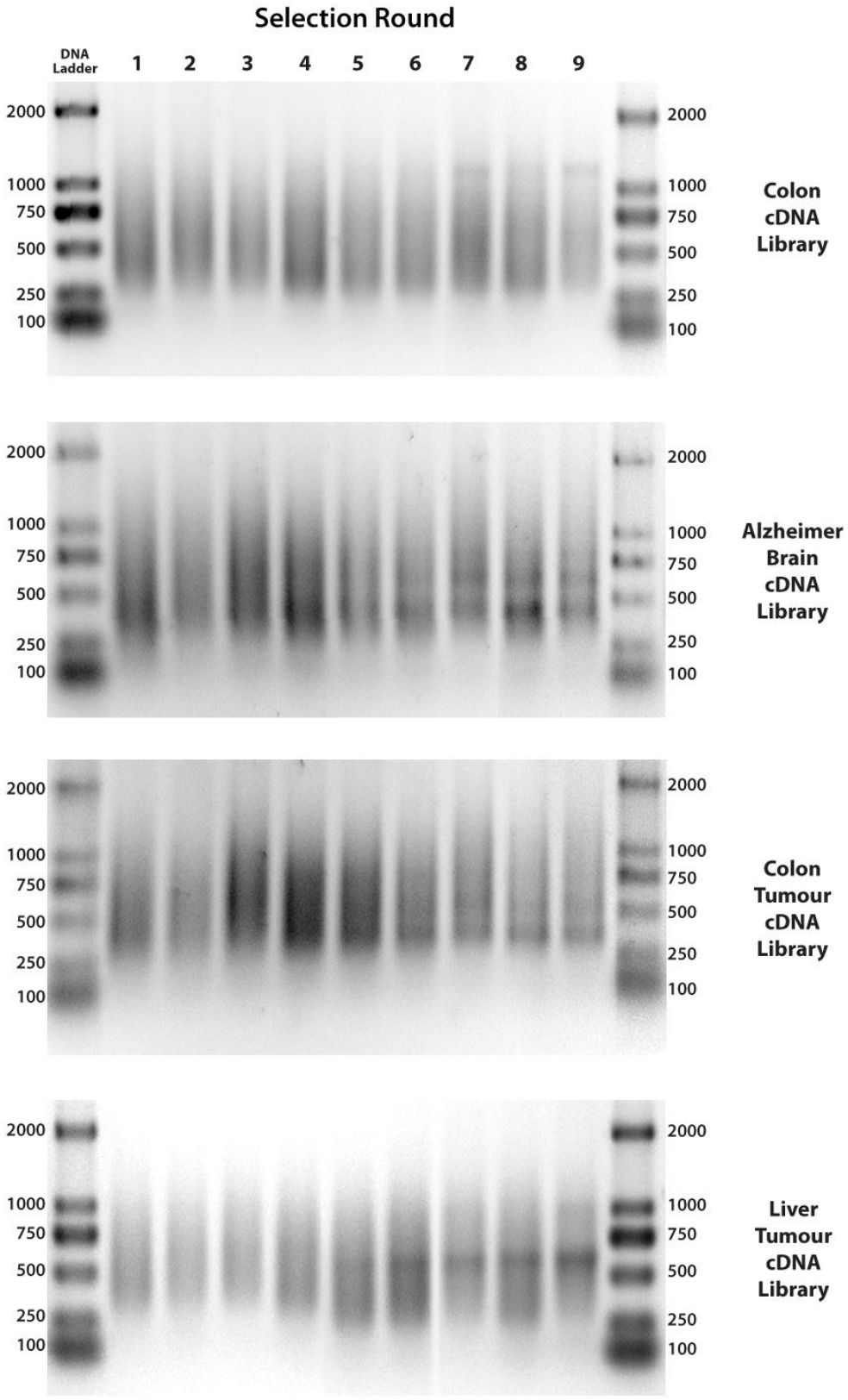


Figure 43: Agarose gel electrophoresis of phage DNA inserts amplified by PCR from colon, Alzheimer's brain, colon tumour and liver tumour cDNA libraries after nine rounds of selection with biotinylated 8-hydroxymanzamine immobilsed on a neutravidin-coated microtitre plate.

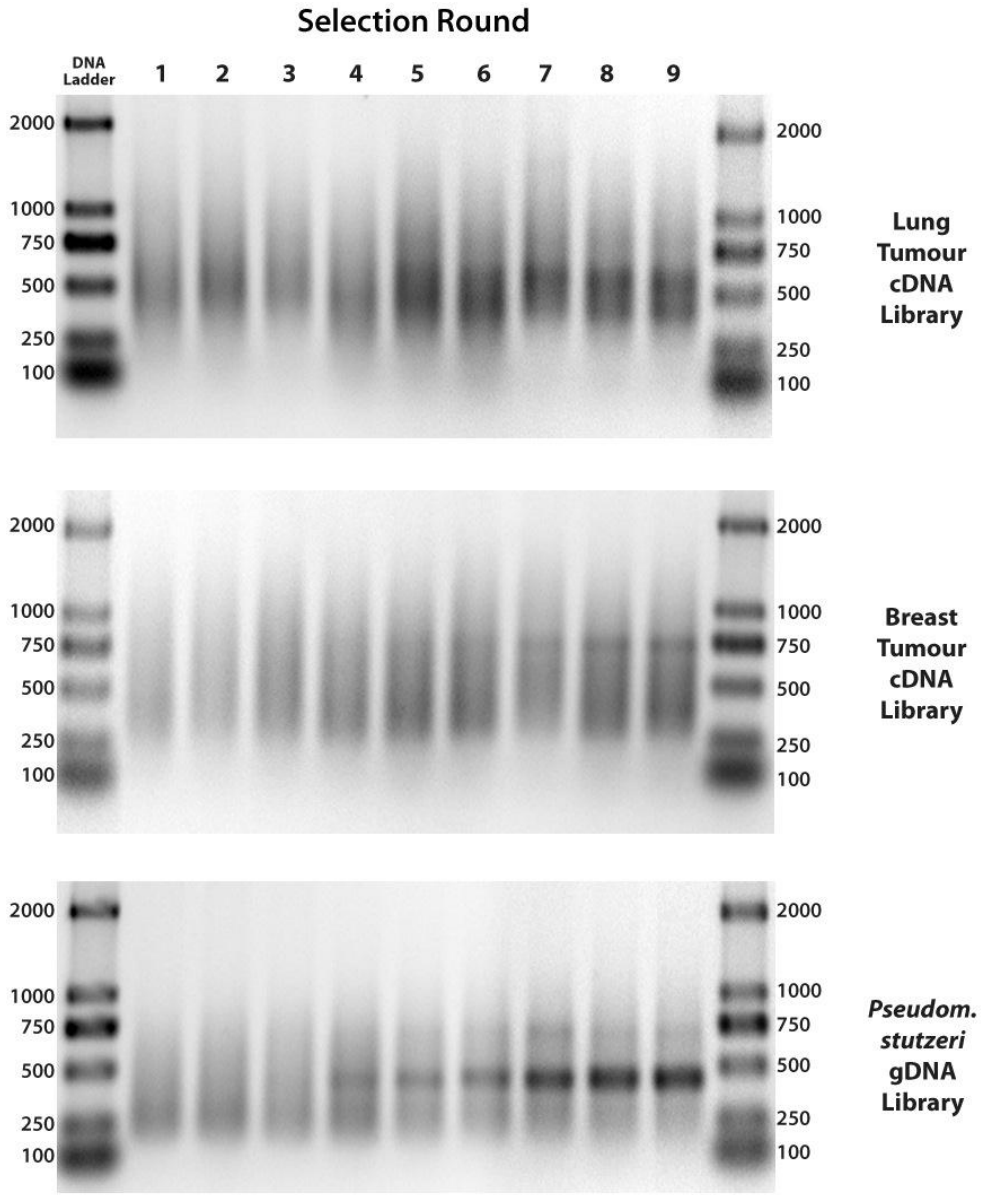


Figure 44: Agarose gel electrophoresis of phage DNA inserts amplified by PCR from lung tumour and breast tumour cDNA libraries as well as *Pseudomonas stutzeri* gDNA library after nine rounds of selection with biotinylated 8-hydroxymanzamine immobilsed on a neutravidin-coated microtitre plate.

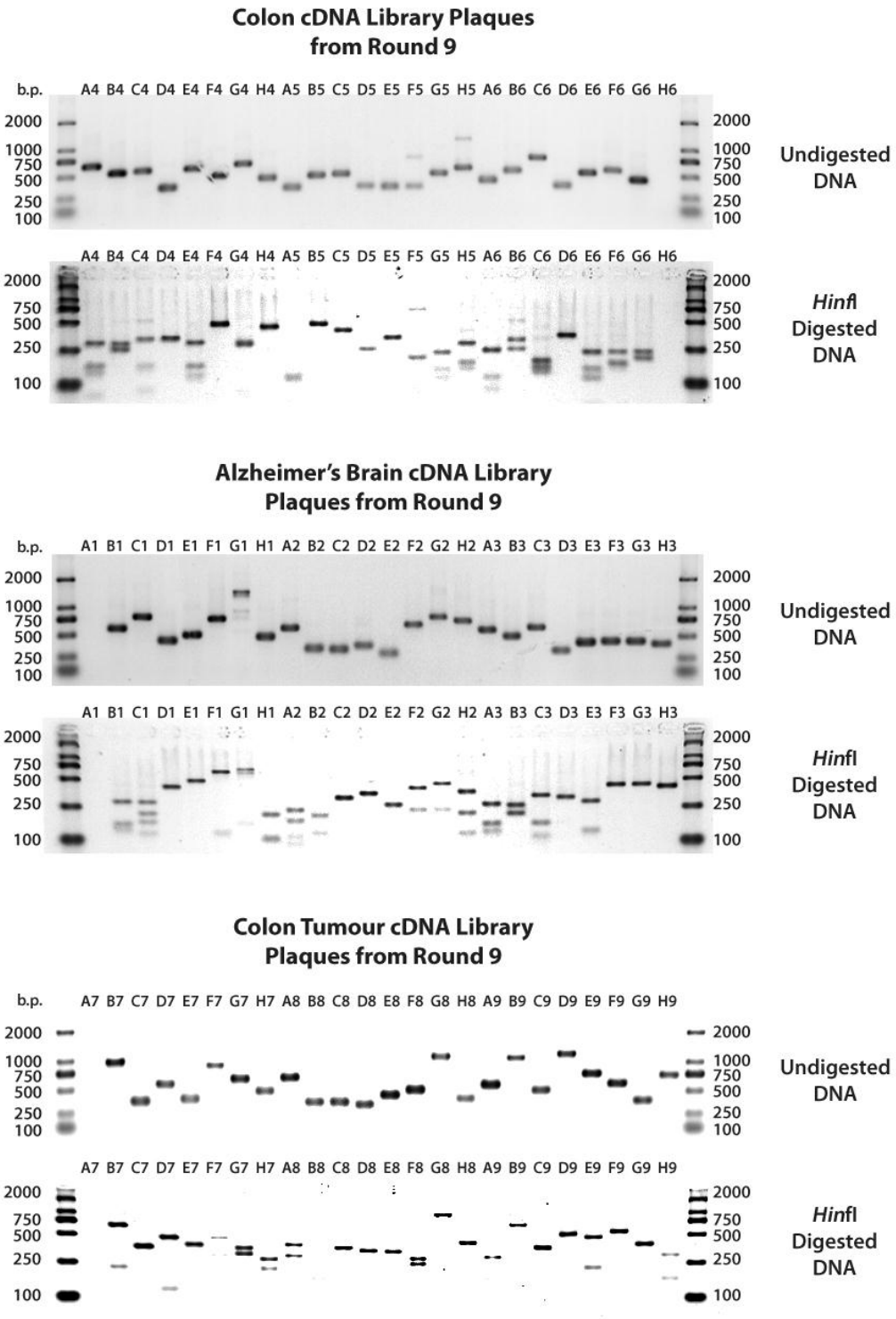


Figure 45: Agarose gel electrophoresis of PCR products obtained from colon, Alzheimer's brain and colon tumour individual plaques after nine rounds of selection with biotinylated 8-hydroxymanzamine immobilised on a neutravidin-coated plate. The DNA inserts, which were amplified using generic T7 primers, were also digested with *Hin*fI to produce unique DNA fingerprints of each clone.

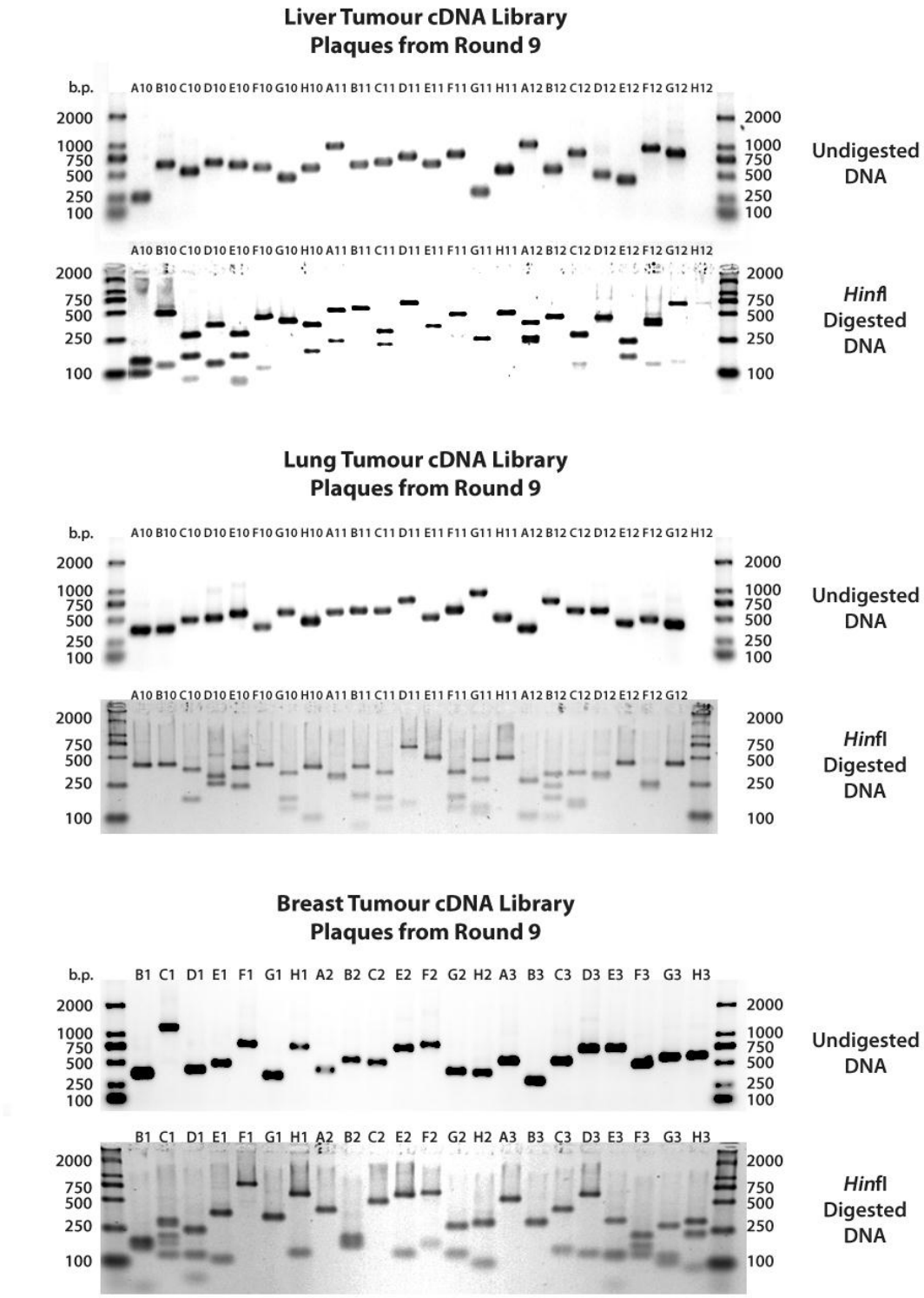


Figure 46: Agarose gel electrophoresis of PCR products obtained from liver tumour, lung tumour and breast tumour individual plaques after nine rounds of selection with biotinylated 8-hydroxymanzamine immobilised on a neutravidin-coated plate. The DNA inserts, which were amplified using generic T7 primers, were also digested with *Hin*fI to produce unique DNA fingerprints of each clone.

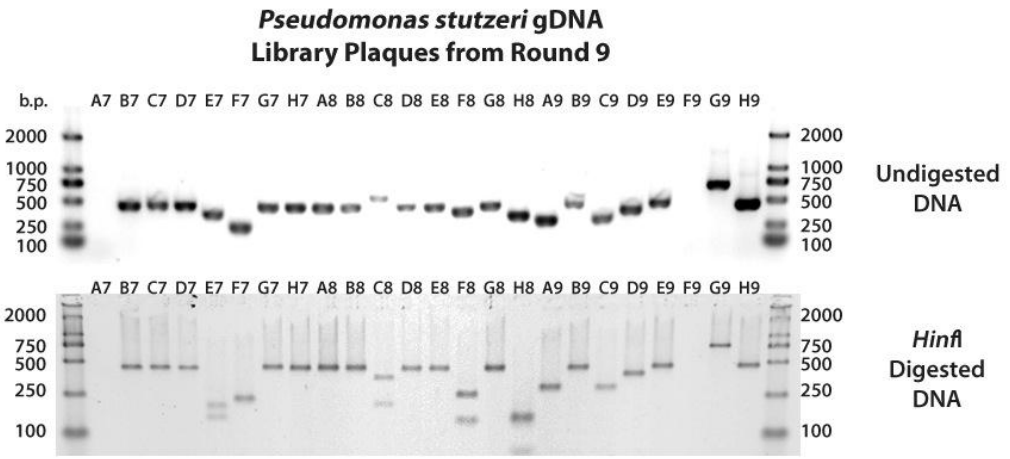


Figure 47: Agarose gel electrophoresis of PCR products obtained from *Pseudomonas stutzeri* individual plaques after nine rounds of selection with biotinylated 8-hydroxymanzamine immobilised on a neutravidin-coated plate. The DNA inserts, which were amplified using generic T7 primers, were also digested with *Hin*fI to produce unique DNA fingerprints of each clone.

Plaques were randomly picked from each round-9 sublibrary, amplified by PCR, digested with *Hin*fI and separated by gel electrophoresis (see Figure 45 - Figure 47). The PCR products from clones containing very similar or identical DNA inserts (as determined by DNA fingerprinting), as well as large DNA inserts, were purified and sequenced (see Table 18).

Initially, the most promising result emerged from the *Pseudomonas stutzeri* gDNA library, which seemed be converging to one clone (~ 500 bp). DNA fingerprinting of the single plaques resulted in one dominant clone (13/23), sized just below 500 bp. Unfortunately, the two representatives sent for DNA sequencing (B7 and H7) showed a peptidyl-prolyl cis-trans isomerase (FKBP-type) gene, which was inserted backwards into the phage DNA. Translating the DNA in frame with the viral coat protein established that a 39 aa long protein is displayed. A protein BLAST identified hypothetical protein PST_1149 as the best match, with a query coverage of 89% and an Expect value of 10^{-3}, but only a maximum identity of 44%. As the displayed peptide is quite long but does not really match any known protein, it is an interesting find. Obviously manzamine has an affinity for this sequence but how this relates to its biological activity in bacteria cannot be determined from this experiment.

In the selections with colon tumour, liver tumour and breast tumour libraries, the round-9 sublibraries appeared as mixtures of clones on the agarose gels. The electrophoresis results of both undigested and *Hin*fI digested DNA fragments isolated from the single plaques displayed no single clone as dominant in any of the libraries.

Table 18: DNA sequencing of PCR products obtained from individual plaques after nine rounds of selection with biotin-manzamine immobilised on a neutravidin-coated PS microtitre plate.

Library	Plaques	Gene Identified from DNA Sequence	Notes	Frame
Col, AB, LuT	-	*no convergence*	-	-
CoT	H7	Homo sap. Myosin XIX (MYO19) transcript var 1+2	phage protein gives dozens of matches, only short query coverage	-
CoT	A8	Serine protease 23 (PRSS23)	late part of CDS	2
CoT	G8	hypothetical LOC100506548, non-coding RNA	gene in backwards	-
CoT	B9	E74-like factor 3 (ETS domain transcription factor, epithelial specific, ELF3)	last 230 of 1120bp CDS	1
LiT	B10	superoxidase dismutase 2, mitochondrial (SOD2), nuclear gene encoding mitochondrial protein	last 245bp of 667bp CDS	-
LiT	C10, E10	ribosomal protein S10	390/320 of 500bp CDS	1
LiT	D10	NIMA (never in mitosis gene a)-related kinase 2 (NEK2)	first 420 of 1150bp CDS	1 or 2
LiT	H10	*no significant similarity*	-	-
LiT	A11	ankyrin repeat domain 10 (ANKR10)	last 300 of 1160 bp CDS	2
LiT	D11	hydroxylacyl-CoA dehydrogenase/3-ketoacyl-CoAthiolase/enoyl-CoA hydratase (trifunctional protein), alpha subunit (HADHA)	gene in backwards	-
BrT	C1, F1, C3, E3, F3	*no significant similarity*	-	-
BrT	E1	phosphohistidine phosphatase 1 (PHPT1)	last 180 of 380bp CDS	?
BrT	H1, D3	ubinuclein 1 (UBN 1)	600 of 3400bp CDS	1
BrT	F2	ribosomal protein L13a (RPL13A)	607 of 612 bp CDS	1
BrT	H2	GLI-Kruppel zinc finger family member (HKR1)	gene in backwards	-
Ps	B7, H7	peptidyl-prolyl cis-trans isomerase, FKBP-type	gene in backwards	-3, -1
Ps	G9	putrescine ABC transporter ATP-binding protein	after other short sequence and stop codon	1

Nevertheless a few clones of each round-9 sublibrary were selected for DNA

sequencing, if there were at least two representatives present. Clone B9 of the colon tumour library was analysed because of its unusually large size of more than 1 kb and resulted in a match for E74-like factor 3 (ETS domain transcription factor, epithelial specific, ELF3). The phage-displayed protein is in frame and covers the last 230 bp of the 1.1 kb CDS protein.

The ETS-domain family of transcription factors is unique to metazoa and one of the largest protein families. The ETS family members are identified through a highly conserved DNA binding domain. The structures of the ETS-domain incorporating proteins are very diverse and no unifying theme has been found regarding their functions. Various subfamilies of ETS-domain proteins are linked to different processes in both the adult and the embryonic development; the promotion of differentiation; the inhibition of apoptosis; the regulation of senescence; the regulation of proliferation in response to ERK signalling, or of haematopoiesis. The development of particular types of cancer is supposedly dependent on ETS-domain proteins to be deregulated by inappropriate expression or expression as fusions with other proteins[368].

If manzamine was to bind specifically to the ETS-domain, this could explain some of the reported bioactivities of the alkaloid. As described in Section 3.2.3.1.1, manzamine A displays antitumour activity against various cancer types like colon tumour, lung carcinoma and breast cancer[269]. Interestingly, colon cancer is one of multiple cancer types exhibiting aberrant GSK-3β protein expression levels[369]. The transcriptional regulation of GSK-3β leading to its abnormal expression in cancers is not known, but Zhang *et al*. recently identified an active Ras – mitogen-activated protein kinase (MAPK) – ETS-2 – p300 cascade leading to GSK-3β overexpression in pancreatic cancer cells[370]. It is important to point out that Ras regulates MAPK through two highly conserved ETS binding elements. Interference with the ETS-domain could impact this cascade and eventually alter the downstream GSK-3β overexpression.

As intriguing as this finding is, clone B9 was the only one ETS-domain was isolated and the expressed protein only represents about 20% of the entire ETS-domain. Further rounds of selection could tell if this or other ETS-domains are enriched, supporting the significance of this interaction.

From the liver tumour library, only two of the analysed clones (C10, E10) had the alien cDNA in frame with the T7 coat protein, both expressing a gene related to ribosomal protein S10 (RPS10). The ambiguous sequencing of clone D10 did not allow us to determine if the gene was in frame. To the best of our knowledge, there are no reports of any interaction between manzamine and RPS10, nor is there any apparent correlation between the compound's bioactivity and the function of RPS10. Considering that ribosomal proteins are highly charged, an interaction with manzamine could simply be based on electrostatic forces. It is not possible to draw any conclusions at this point - further rounds of selection are required to allow the sublibraries to converge.

The selection rounds performed with the breast tumour library did not generate a dominating clone either. However, two of the isolated plaques (H1 and F2) encoded CDS-conatined proteins in frame with the T7 coat protein. Clone H1 incorporated an approximate 600 bp part of the 3.4 kb CDS of ubinuclein 1 (UBN-1), or ubiquitously expressed nuclear protein. This protein regulates cell senescence and is involved in the formation of senescence-associated heterochromatin foci (SAHF). Surprisingly, the cDNA insert present in the phage has part of the gene deleted after 70 bp, resulting in a frame shift and the remaining gene is out of frame. The displayed protein is 62 aa long and only the first 21 aa represent UBN-1 protein.

Clone F2 contained in frame ribosomal protein L13a (RPL13a), displaying 200 of 203 aa. Again, there is no evidence that manzamine and RPL13a interact nor can we see any correlation between the compound's bioactivity and the function of RPL13a or the ribosome.

The colon, Alzheimer's brain and lung tumour libraries were insufficiently converged for sequencing. A few additional rounds of selection seem to be required to verify if the libraries will converge at all.

The biopanning of biotin-manzamine with all previously tested libraries was hence continued up to round-15 and carried out similarly as previously described (for figures displaying agarose gels of sublibraries, single plaques and enzymatic digest of picked plaques, see Section 8.3.2.1). Selected clones were submitted for sequencing and the results are presented in Table 19.

Table 19: DNA sequencing of PCR products obtained from individual plaques after 15 rounds of selection with biotin-manzamine immobilised on a neutravidin-coated PS microtitre plate.

Library	Plaques	Gene Identified from DNA Sequence	Notes	Frame
AB	A1	ribosomal protein L13a	610bp of 612bp CDS	1
AB	B3	protein kinase C, beta, transcript variant 2	gene in backwards	-
AB	D2	SRY (sex determining region Y)-box 6 (SOX6), RefSeqGene on chromosome 11	-	
Col	C4	ATP synthase, H+ transporting, mitochondrial F_0 complex, subunit C3 (subunit 9) (ATP5G3), nuclear gene encoding mitochondrial protein	270bp of 530bp CDS	1
Col	D4	PRELI domain containing 1 (PRELID1), nuclear gene encoding mitochondrial protein	310bp of 660bp CDS	1
Col	F4	acyl-CoA thioesterase 11 (ACOT11), transcript variant 2	gene in backwards	-
Col	H4	GRB10 interacting GYF protein 1 (GIGYF1), mRNA	gene in backwards	-
CoT	B7	ribosomal protein L22	170bp of 390bp CDS	1
CoT	C7	apolipoprotein L, 1 (APOL1), transcript variant 4	230bp of 1140bp CDS	1
CoT	G8	THAP domain containing 5 (THAP5), transcript variant 1	230bp of 1190bp CDS	1
LiT	A11	coiled-coil domain containing 12 (CCDC12)	230bp of 540bp CDS	1
LiT	B10	zinc finger protein 398 (ZNF398), transcript variant 2	outside CDS	-
LiT	D10	G antigen 2C (GAGE2C)	270bp of 350bp CDS	1
LiT	D11	annexin A2 (ANXA2), transcript variant 3	350bp of 1020bp CDS	1
LiT	D12	poly(rC) binding protein 2 (PCBP2), transcript variant 1	515bp of 1100bp CDS	1
LiT	E10	keratin 8 (KRT8), transcript variant 3	last 300bp of 1450bp CDS	1
LiT	H11	UDP-N-acetyl-alpha-D-galactosamine: poly-peptide N-acetylgalactosaminyltransferase 6 (GalNAc-T6) (GALNT6)	outside CDS	-

* Plaques derived from this library were picked after 12 rounds of selection

(Table 17 continued)

Library	Plaques	Gene Identified from DNA Sequence	Notes	Frame
LuT	C3	filamin B, beta (FLNB), transcript variant 4	outside CDS	-
LuT	E1	KIAA0368 (KIAA0368)	720bp of 6000bp CDS	1
LuT	G2	U2 snRNP-associated SURP domain containing (U2SURP), mRNA	gene in backwards	-
BrT	A5	5'-nucleotidase domain containing 3 (NT5DC3)	outside CDS	-
BrT	B5	lysocardiolipin acyltransferase 1 (LCLAT1), transcript variant 1	gene in backwards	-
BrT	D4	rhomboid 5 homolog 1 (Drosophila) (RHBDF1)	230bp of 2570bp CDS	3
BrT	G5	NIMA (never in mitosis gene a)-related kinase 4 (NEK4), transcript variant 1	190bp of 2520bp CDS	2
BrT	H5	teashirt zinc finger homeobox 2 (TSHZ2), transcript variant 2	outside CDS	-
Ps	C7	peptidyl-prolyl cis-trans isomerase, FKBP-type & hypothetical protein	gene in backwards	-
Ps	E8	T7 Flexi Vector pFC30K His6HaloTag, complete sequence	Not a *Pseudomonas* gene	-
Ps	F7	compare C7, but slightly different clone	gene in backwards	-
Ps	H7	identical to clone F7	gene in backwards	-

No consensus was reached in any of the sublibraries of round-15. Ten of the sequenced clones had DNA inserts which resulted in T7 phage-displayed human protein CDS sequences or parts thereof (AB_A1, Col_C4, Col_D4, CoT_B7, CoT_C7, CoT_G8, LiT_A11, LiT_D10, LiT_E10, LuT_E1). However, none of these clones dominated any single, or occurred in multiple libraries. The majority of analysed clones seemed insignificant, i.e. the cDNA insert was in backward, out of frame, or outside the CDS. Clone AB_A1 appeared as the only possible real hit, as it displayed almost the entire CDS sequence of ribosomal protein L13. Yet, the low abundance of this clone throughout the entire data set, suggested that the biopanning has failed for biotin-manzamine. This conclusion is supported by the isolation of cloning vectors and the high number of clones with genes inserted backwards.

4.4 Daptomycin

4.4.1 Introduction

Daptomycin is a branched, cyclic lipopeptide antibiotic used to treat skin infections caused by Gram-positive bacteria[280]. Its characteristics and the current opinion on the antibacterial mode of action are reviewed in Section 3.2.3.2.1, which outlines the limited knowledge on its precise mode of action and lack of knowledge about human off-targets.

Reverse chemical proteomics is the ideal tool to help elucidate targets and off-targets for this complex compound as it allows for rapid discovery of cognate drug-receptor pairs. As so little is known about off-targets or other potential medical applications beyond its antibiotic capacity, we utilise a wide selection of cDNA libraries derived from both normal and diseased cells of various tissues. Our selection of DNA libraries causes a strong focus on isolating human cellular off-targets, but also includes one gDNA library of a Gram-negative bacterium. Daptomycin shows no bactericidal or bacteriostatic effect on Gram-negative organisms but in the context of phage-displayed libraries, this is not really relevant to how and if a drug can enter the cytoplasm of a Gram-positive organism..

4.4.2 Results and discussion

A biotinylated analogue of daptomycin, containing a long, hydrophilic TEG linker was synthesised (**101**, see Scheme 5) and immobilised on a neutravidin-coated PS microtitre plate to generate affinity support for display cloning. A similar biotinylated compound (**103**), containing only the ornithine side chain of daptomycin at the end of the linker, was synthesised and immobilised on a second neutravidin-coated plate as a control.

Similarly to Section 4.3, six T7 phage-displayed human cDNA libraries (colon; Alzheimer's brain; cancerous cells of breast, colon, liver and lung tissue) and the gDNA library of the Gram-negative *Pseudomonas stutzeri* were then subjected to nine rounds of selection using a daptomycin carrying affinity support for biopanning. Non-specific binders were cleared by preincubation with a control probe containing only the ornithine sidechain.

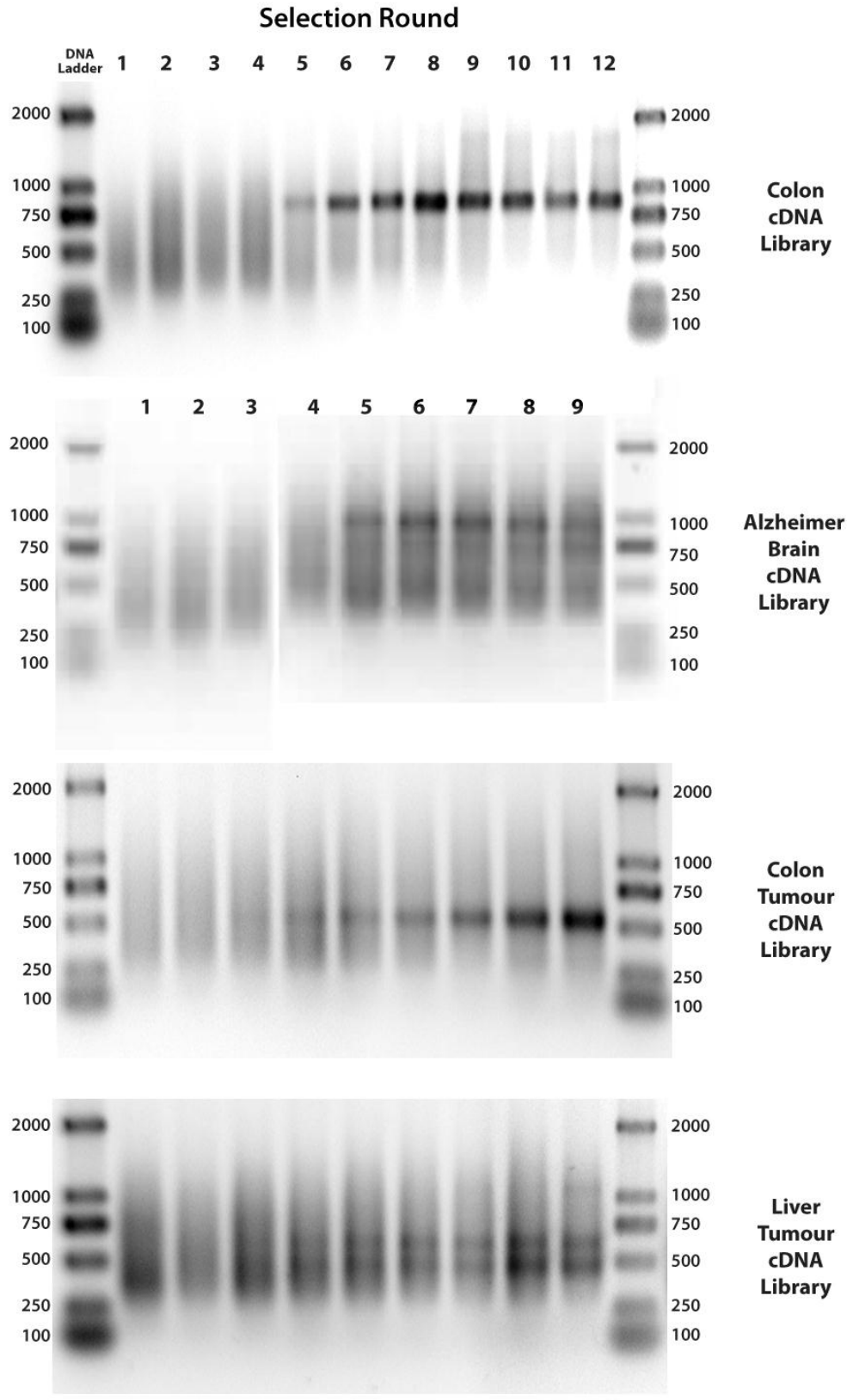


Figure 48: Agarose gel electrophoresis of phage DNA inserts amplified by PCR from colon, Alzheimer's brain, colon tumour and liver tumour cDNA libraries after twelve or nine rounds of selection with biotinylated daptomycin immobilised on a neutravidin-coated microtitre plate.

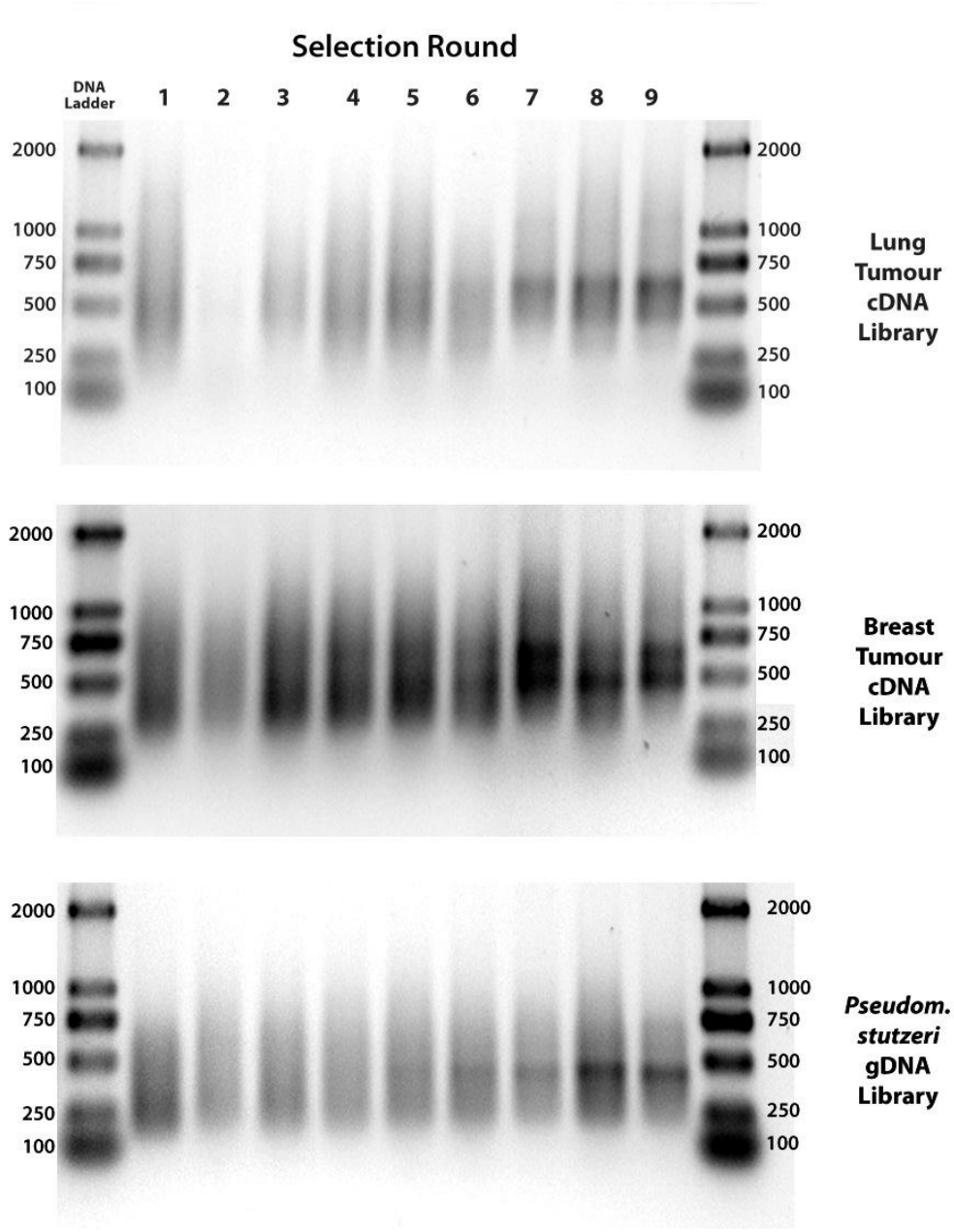


Figure 49: Agarose gel electrophoresis of phage DNA inserts amplified by PCR from lung tumour and breast tumour cDNA, and *Pseudomonas stutzeri* gDNA-libraries after nine rounds of selection with biotinylated daptomycin immobilised on a neutravidin-coated microtitre plate

The selections were carried out in identical manner to the biotinylated manzamine selection described in Section 4.3. If convergence occurred within the nine rounds of selection (see Figure 48 and Figure 49), 23 plaques were picked randomly from each round-9 sublibrary.

For the colon library, three additional rounds of selection were performed to achieve complete convergence and remove any non-specific binders. In this case 24 plaques were randomly selected from the round-12 sublibrary. The plaques' DNA inserts were amplified by PCR, digested with *Hin*fI and separated by gel electrophoresis (see Figure 50 and Figure 51). The PCR products from clones containing very similar or

identical DNA inserts (as determined by DNA fingerprinting), as well as large DNA inserts, were purified and sequenced (see Table 20).

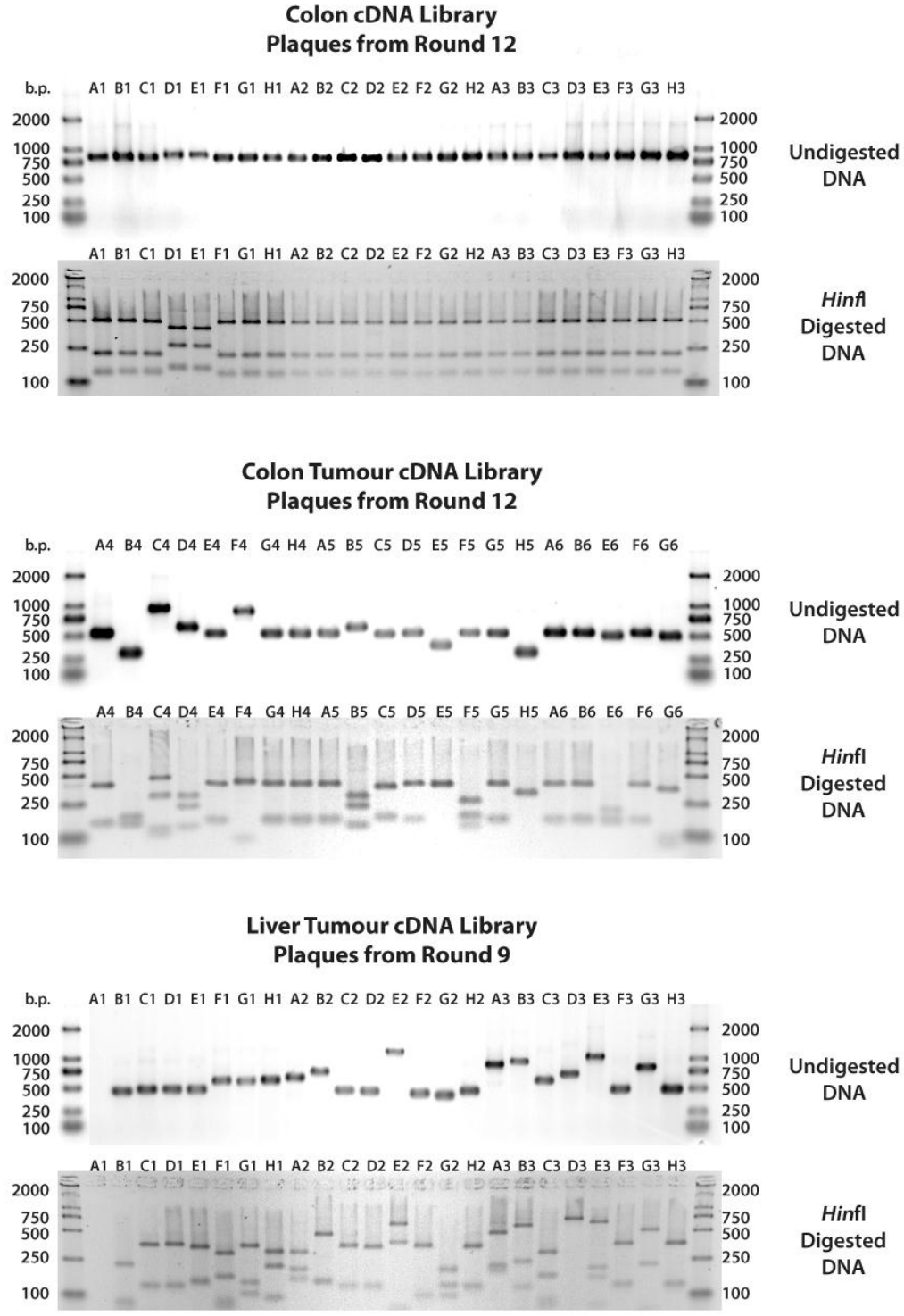


Figure 50: Agarose gel electrophoresis of PCR products obtained from colon, colon tumour and liver tumour individual plaques after twelve or nine rounds of selection with biotinylated daptomycin immobilised on a neutravidin-coated plate. The DNA inserts, which were amplified using generic T7 primers, were also digested with *Hin*fI to produce unique DNA fingerprints of each clone.

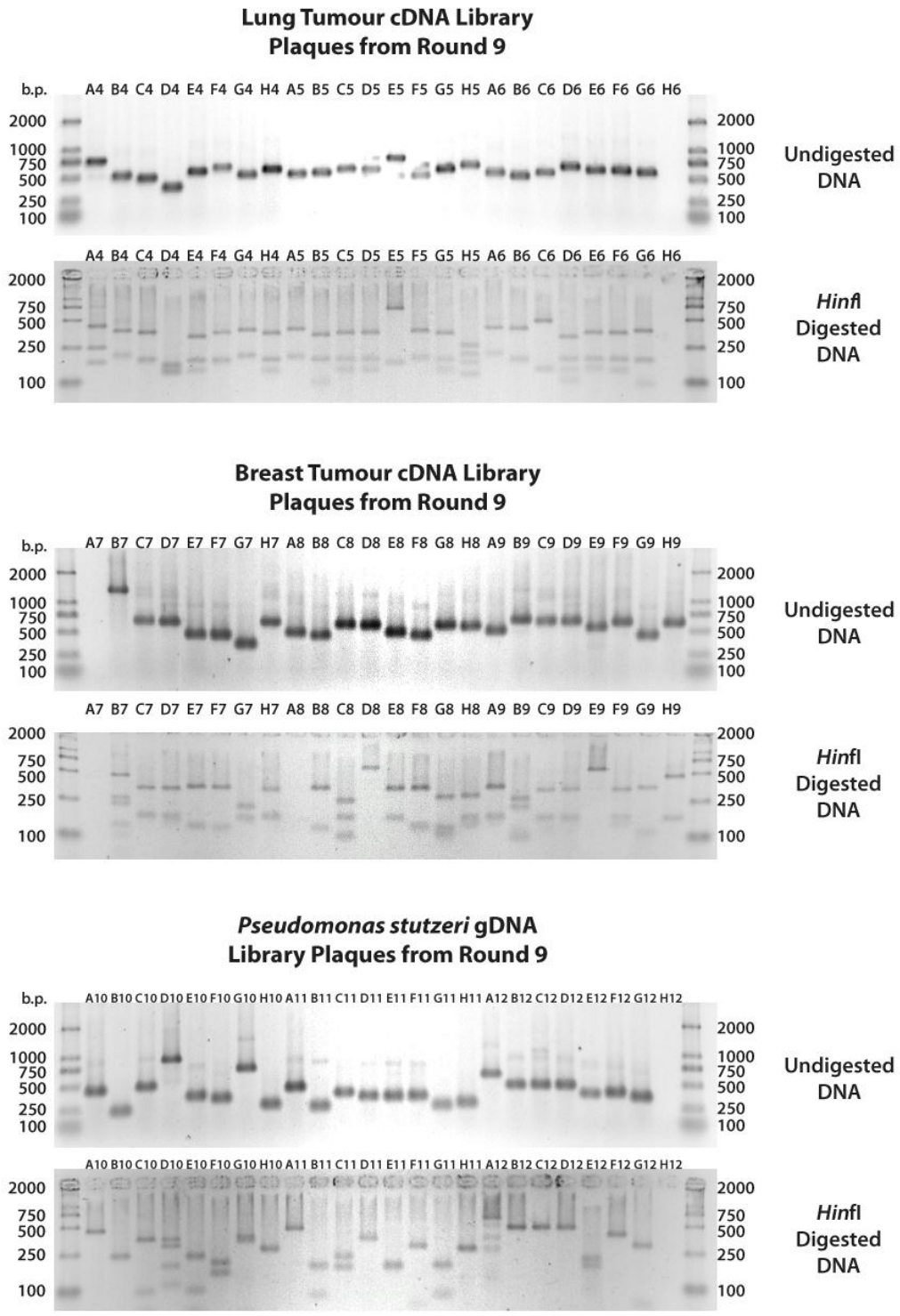


Figure 51: Agarose gel electrophoresis of PCR products obtained from lung tumour, breast tumour and *P. stutzeri* individual plaques after nine rounds of selection with biotinylated daptomycin immobilised on a neutravidin-coated plate. The DNA inserts, which were amplified using generic T7 primers, were also digested with *Hin*fI to produce unique DNA fingerprints of each clone.

Table 20: DNA sequencing of PCR products obtained from individaul plaques after nine rounds of selection with biotin-daptomycin immobilised on a neutravidin-coated PS microtitre plate.

Library	Plaques	Gene Identified from DNA Sequence	Notes	Frame
Col*	A3	leukocyte receptor cluster (LRC) member 8	minor fraction of CDS	1
Col*	A3	kinesin family member 1C	after stop codon	-
Col*	D1	family with sequence similarity 114, member A2	early 250bp of 1500bp CDS	1
Col*	D1	tenascin C (TNC)	500bp of 6400bp CDS	1
AB	-	*no convergence*	-	-
CoT	A4, E4, G4, H4	**ribosomal protein S19**	350bp of 420bp CDS	1
CoT	B4	dynamin 2 (DNM2),	only late 180bp of 2500bp CDS	2
CoT	B5	sphingomyelin phosphodiesterase 4	outside CDS	-
LiT	C1, D1, C2, H3	**ribosomal protein S19**	350bp of 420bp CDS	1
LiT	E1	**ribosomal protein S19**	360bp of 420bp CDS	1
LiT	B2	required for meiotic nuclear division 5 homolog B	gene in backwards	-
LiT	A3	UDP-Gal:betaGlcNAc beta 1,4-galactosyltransferase, polypeptide 5	gene in backwards	-
LiT	B3	RAD21 homolog (S. pombe)	500bp of 1800bp CDS	1
LuT	A4, C5	ribosomal protein S10	last 360bp of 500bp CDS	1
LuT	C4	**ribosomal protein S19**	300bp of 420bp CDS	1
BrT	B7	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro	very late 480bp of 3680bp CDS	1
BrT	D7	**ribosomal protein S19**	300bp of 420bp CDS	1
BrT	A9	**ribosomal protein S19**	330bp of 420bp CDS	1
BrT	D9	ribosomal protein S10	430bp of 500bp CDS	1
BrT	E9	multiple matches only between 200-250bp		-
BrT	H9	LYR motif containing 5	230bp of 270bp CDS	1
Ps	C10	flagellar biosynthesis protein	-	-2
Ps	D10	*no significant similarities*	-	-
Ps	F10	tRNA 2-selenouridine synthase	-	1
Ps	B11	thymidilate kinase	-	3

* Plaques derived from this library were picked after 12 rounds of selection

Analysis of the DNA sequences from a subset of the rescued phages (see Section 8.3.3.3) clearly showed that the clones displaying the ribosomal protein S19 (RPS19) were the most abundant in most libraries. They comprised 11 out of 23 plaques selected from the colon tumour library, 8 out of 23 plaques from the liver tumour library, 7 out of 23 from the lung tumour library, and 5 out of 23 from the breast tumour library. In conjunction with an exponential rise in titre, this indicates a successful selection. The probability of isolating two clones expressing the same protein from within the same library is $1:10^{7}$. To isolate different clones expressing the same protein from four different libraries, the probability is $1:10^{21}$.

```
                    10         20         30         40         50
                ....|....| ....|....| ....|....| ....|....| ....|....|
CoT_A4      :   ---------- ---------- -SKKSGKLKV PEWVDTVKLA KHKELAPYDE
LiT_C1      :   ---------- ---------- ---------- --SVDTVKLA KHKELAPYDE
LiT_E1      :   ---------- ---------- ---------- --------SA KHKELAPYDE
LuT_C4      :   ---------- ---------- ---------- -----SVKLA KHKELAPYDE
BrT_D7      :   ---------- ---------- ---------- -----GVKLA KHKELAPYDE
BrT_A9      :   ---------- ---------- ---------- --SVDTVKLA KHKELAPYDE
RPS19       :   MPGVTVKDVN QQEFVRALAA FLKKSGKLKV PEWVDTVKLA KHKELAPYDE
Consensus   :   ---------- ---------- ---------- ---------A KHKELAPYDE

                    60         70         80         90        100
                ....|....| ....|....| ....|....| ....|....| ....|....|
CoT_A4      :   NWFYTRAAST ARHLYLRGGA GVGSMTKIYG GRQRNGVMPS HFSRGSKSVA
LiT_C1      :   NWFYTRAAST ARHLYLRGGA GVGSMTKIYG GRQRNGVMPS HFSRGSKSVA
LiT_E1      :   NWFYTRAAST ARHLYLRGGA GVGSMTKIYG GRQRNGVMPS HFSRGSKSVA
LuT_C4      :   NWFYTRAAST ARHLYLRGGA GVGSMTKIYG GRQRNGVMPS HFSRGSKSVA
BrT_D7      :   NWFYTRAAST ARHLYLRGGA GVGSMTKIYG GRQRNGVMPS HFSRGSKSVA
BrT_A9      :   NWFYTRAAST ARHLYLRGGA GVGSMTKIYG GRQRNGVMPS HFSRGSKSVA
RPS19       :   NWFYTRAAST ARHLYLRGGA GVGSMTKIYG GRQRNGVMPS HFSRGSKSVA
Consensus   :   NWFYTRAAST ARHLYLRGGA GVGSMTKIYG GRQRNGVMPS HFSRGSKSVA

                   110        120        130        140
                ....|....| ....|....| ....|....| ....|....| ....|
CoT_A4      :   RRVLQALEGL KMVEKDQDGG RKLTPQGQRD LDRIAGQVAA ANKKH
LiT_C1      :   RRVLQALEGL KMVEKDQDGG RKLTPQGQRD LDRIAGQVAA ANKKH
LiT_E1      :   RRVLQALEGL KMVEKDQDGG RKLTPQGQRD LDRIAGQVAA ANKKH
LuT_C4      :   RRVLQALEGL KMVEKDQDGG RKLTPQGQRD LDRIAGQVAA ANKKH
BrT_D7      :   RRVLQALEGL KMVEKDQDGG RKLTPQGQRD LDRIAGQVAA ANKKH
BrT_A9      :   RRVLQALEGL KMVEKDQDGG RKLTPQGQRD LDRIAGQVAA ANKKH
RPS19       :   RRVLQALEGL KMVEKDQDGG RKLTPQGQRD LDRIAGQVAA ANKKH
Consensus   :   RRVLQALEGL KMVEKDQDGG RKLTPQGQRD LDRIAGQVAA ANKKH
```

Figure 52: Alignment of T7 phage-displayed RPS19 protein sequences isolated from colon tumour, liver tumour, lung tumour and breast tumour libraries with the human RPS19 analogue (NCBI ID: NP_001013.1)

Alignment of all converted RPS19 proteins with the authentic human RPS19 (see Figure 52) revealed that all the rescued clones were missing 21-39 N-terminal amino acids with various lengths of 3'-UTR. The longest protein was from the colon tumour library (A4), which lacks only the first 21 aa from the RPS19 protein. In all other cases (LiTC1, LiTE1, LuTC4, BrTD7, BrTA9), at least 106 out of 145 aa of the human RPS19 were present, suggesting that the binding site for daptomycin was after the first 39 residues.

Recently the total structure of the eukaryotic ribosome has been published and shows RPS19 to be located on the outside of the ribosome[371]. As displayed through structural and functional analysis by Gregory *et al.*[372], RPS19 inherits two highly basic patches (residues 52-62 and 101-121), both attributed as playing an essential role in RPS19's capacity to support cell growth.

The RPS19 protein is a component of the 40S ribosomal subunit and belongs to a family of ribosomal proteins restricted to eukaryotes and archaea. There is no homologue of RPS19 in bacteria. It is essential for yeast viability and for early stages of development in mice[373,374]. Disruption as well as point mutations of the RPS19 gene in yeast and human cells affect maturation of the pre-ribosomal RNA (pre-rRNA) and block production of the 40S ribosomal subunits[373,375,376]. Mutations of RPS19, as well as of two other ribosomal proteins, RPS24 and RPS17, have been linked to the rare congenital disease Diamond-Blackfan Anemia (DBA) [376-380]. DBA is characterised by the defective differentiation of pro-erythroblasts, the precursors of red blood cells. Patients suffer severe anemia and display heterogeneous clinical features including malformations, growth failure and predisposition to cancer[381,382]. Why the mutation of a ribosomal protein primarily affects pro-erythroblast differentiation still remains unclear.

Further studies on the function of RPS19 suggest additional extra-ribosomal functions for RPS19. Kondoh *et al.*[383] reported higher expression levels of RPS19 in certain colon cancer cell lines, compared to normal colon tissue, which increased concomitantly with tumour progression in two pairs of cell lines derived from the same patients and decreased after butyrate-treatment in the HT29 cell line. The authors correlated high expression with higher malignant potential of colon carcinoma cells because there was no association between the high levels and the cell growth states.

Table 21: Amino acid composition of RPS19 and its smallest analogue isolated from plaque E1 from liver tumour library using biotin-daptomycin immobilised on neutravidin-coated plates.

	RPS19		Isolated Protein (LiT E1)		
Residue	**Number**	**Mole%**	**Number**	**Mole%**	**missing**
A = Ala	15	10.3	12	11.2	3
D = Asp	7	4.8	5	4.7	2
E = Glu	6	4.1	4	3.7	2
F = Phe	4	2.8	2	1.9	2
G = Gly	15	10.3	13	12.2	2
H = His	4	2.8	4	3.7	0
I = Ile	2	1.4	2	1.9	0
K = Lys	15	10.3	9	8.4	6
L = Leu	12	8.3	8	7.5	4
M = Met	4	2.8	3	2.8	1
N = Asn	4	2.8	3	2.8	1
P = Pro	5	3.4	3	2.8	2
Q = Gln	8	5.5	6	5.6	2
R = Arg	12	8.3	11	10.3	1
S = Ser	7	4.8	7	6.5	0
T = Thr	6	4.1	4	3.7	2
V = Val	13	9.0	6	5.6	7
W = Trp	2	1.4	1	0.9	1
Y = Tyr	4	2.8	4	3.7	0

Property	**Residues**	**Number**	**Mole%**	**Number**	**Mole%**
Aliphatic	(A+I+L+V)	42	29.0	28	26.2
Aromatic	(F+H+W+Y)	14	9.7	11	10.3
Non-polar	(A+F+G+I+L+M+P+V+W+Y)	76	52.4	54	50.5
Polar	(D+E+H+K+N+Q+R+S+T)	69	47.6	53	49.5
Charged	(D+E+H+K+R)	44	30.3	33	30.8
Basic	(H+K+R)	31	21.4	24	22.4
Acidic	(D+E)	13	9.0	9	8.4

When analysing the amino acid composition of the isolated protein (see Table 21), a high number of charged residues becomes obvious. Within the isolated RPS19

fragments, 22.4% of the residues carry a basic charge and 8.4% are acidic, which is complementary to the overall negative charge of daptomycin.
Daptomycin is known to chelate divalent ions, like calcium or magnesium - a step required for its antibiotic activity. During the biopanning selection, the phage is cultivated in media containing 1 mM Mg^{2+} and 0.45 µM Ca^{2+}, suggesting that daptomycin is at least partially chelated, which may aid binding to charged targets such as RPS19.
An on-phage binding assay was performed to determine if the T7 phage-displayed RPS19 clones (clone C1 from Liver tumour library) isolated had greater affinity for neutravidin-coated plates derivatised with biotinylated daptomycin than for similar plates derivatised with a control compound (see Figure 53). A T7 wild type clone, lacking a cDNA insert, was used in parallel as a negative control. The assay was conducted in triplicate and the results show a statistically significant higher binding to daptomycin derivatised plates. Further validation is required and will be part of future endeavours.

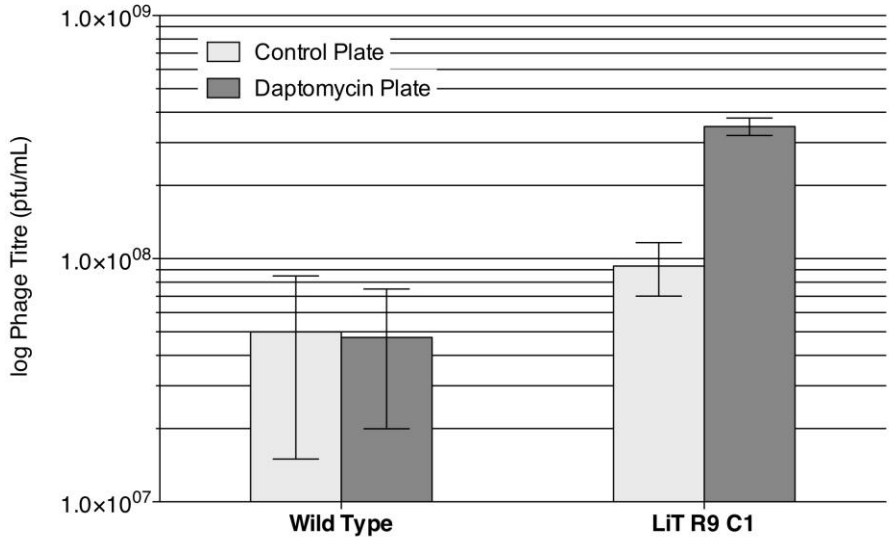


Figure 53: On-phage binding study comparing the affinity of the RPS19-displaying phage clone C1 (from liver tumour sublibrary of round-9) for neutravidin-coated plates derivatised with a control compound and a similar plate derivatised with biotinylated daptomycin.

The round-9 sublibraries from Alzheimer's and *P. stuzeri*, which had not converged, were reamplified to continue the biopanning for another six rounds. Agarose gels and sequences from single plaques picked after round-15 are displayed in Section 8.3.2.2. Selected clones were sent for sequencing and the results are presented in Table 22. No consensus was reached in any of the libraries and the isolated clones did not look like they were significant. Some were the same as previously isolated from round 9

but overall the results suggest that if the library has not converged by round 7 or 8 that it is not going to.

Table 22: DNA sequencing of PCR products obtained from individual plaques after 15 rounds of selection with biotin-daptomycin immobilised on a neutravidin-coated PS microtitre plate

Library	Plaques	Gene Identified from DNA Sequence	Notes	Frame
AB	B4	dystrobrevin, alpha (DTNA), transcript variant 16	last 300bp of 1200bp CDS	1
AB	B5	leukocyte receptor cluster (LRC) member 8	first 500bp of 2800bp CDS	1
AB	C6	chromosome 2 genomic contig, alternate assembly HuRef SCAF_1103279187422	no CDS	-
AB	F4	cell division cycle and apoptosis regulator 1 (CCAR1)	last 230bp of 3450bp CDS	1
AB	H5	serine/arginine-rich splicing factor 2 (SRSF2), non-coding RNA	no CDS	1
Ps	B7	aminopeptidase	gene in backwards	-
Ps	B9	Homo sapiens interleukin 1 receptor accessory protein-like 2 (IL1RAPL2); and testis expressed 13A (TEX13A)	contaminant	-
Ps	C7	transporter	100aa of 670aa protein	1
Ps	H7	peptidyl-prolyl cis-trans isomerase, FKBP-type & hypothetical protein	gene in backwards	-

4.5 Artesunate

4.5.1 Introduction

Artesunate (**106**) is part of the artemisinin group of drugs and is a semisynthetic derivative of the natural product artemisinin (**104**), a highly oxygenated sesquiterpene isolated from *Artemisia annua*[320].

Artemisinin and its derivatives are recommended for combination therapy with other antimalarial drugs by the World Health Organisation (WHO) for treatment of both chloroquine-sensitive and chloroquine-resistant strains of *Plasmodium falciparum*[335]. The unusual structure and therapeutic significance of this compound class resulted in wide-ranging research into its chemistry and pharmacology, not only as an antimalarial but also as a new drug for anticancer treatment[321]. Various members of the artemisinin group showed cytotoxicity in nano- and low micromolar range against breast, colon, lung and ovarian tumour cells as well as central nervous system tumours and leukaemia[322]. The mechanisms of action for these activities are widely discussed and multilayered modes of action have been suggested, including alkylating potential after radical formation upon activation of the endoperoxide bridge[321]. However, out of the myriad of potential cellular targets and of factors involved in the profound cytotoxicity, no human target has been identified[384]. As malaria T7 cDNA libraries do not exist, we attempted to determine only human receptors for artesunate.

4.5.2 Results and discussion

A biotinylated analogue of artesunate, containing a long, hydrophilic TEG linker was synthesised (**123**; see Scheme 9) and immobilised on a neutravidin-coated PS microtitre plate to generate an affinity support for display cloning. A similar biotinylated compound (**124**), incorporating valeroyl amide at the end of the linker in place of artesunate, was synthesised and immobilised on a second neutravidin-coated plate as a control.

T7 phage-displayed human cDNA libraries (colon; Alzheimer's brain; cancerous cells of breast, colon, liver and lung tissue) were then subjected to nine rounds of selection using the artesunate-carrying affinity support for biopanning.

The selections were performed identically to those with biotinylated manzamine described in Section 3.2.3.1.2. If convergence occurred within the nine rounds of selection (see Figure 54 and Figure 55), plaques were picked randomly from each

round-9 sublibrary and *Hin*f1 digested (see Figure 56 and Figure 57). The PCR products from clones that appeared multiple times were purified and sequenced (see Table 23).

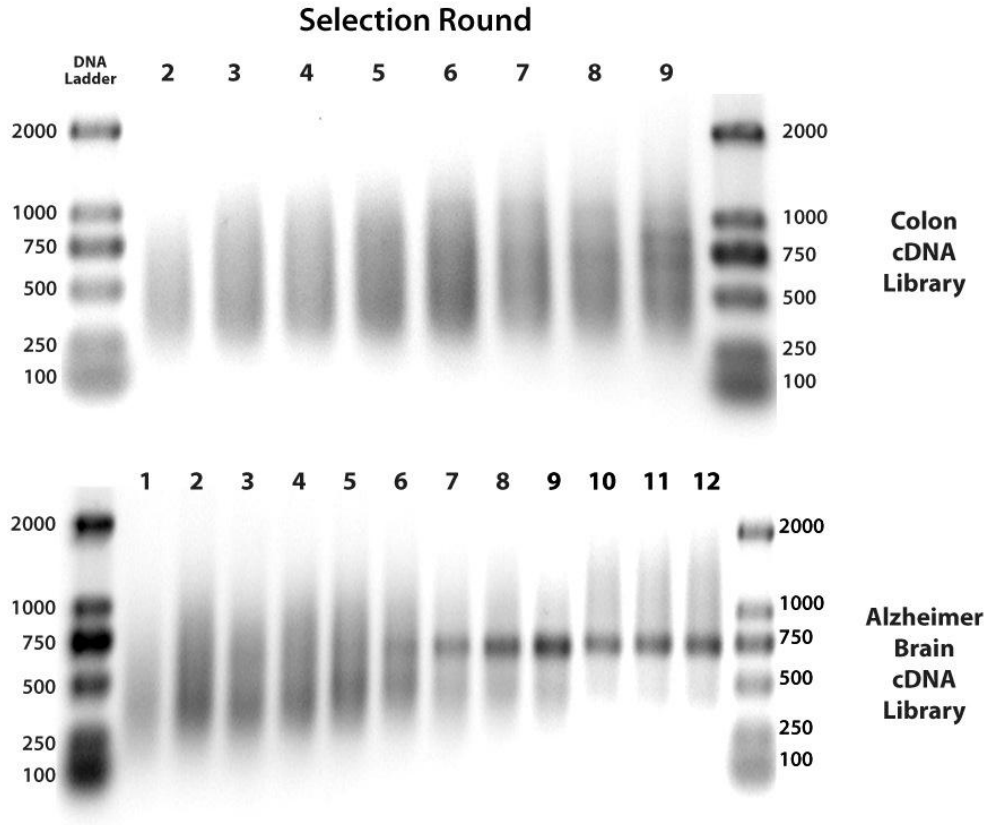


Figure 54: Agarose gel electrophoresis of phage DNA inserts amplified by PCR from colon and Alzheimer's brain cDNA libraries after nine or twelve rounds of selection with biotinylated artesunate immobilised on a neutravidin-coated microtitre plate.

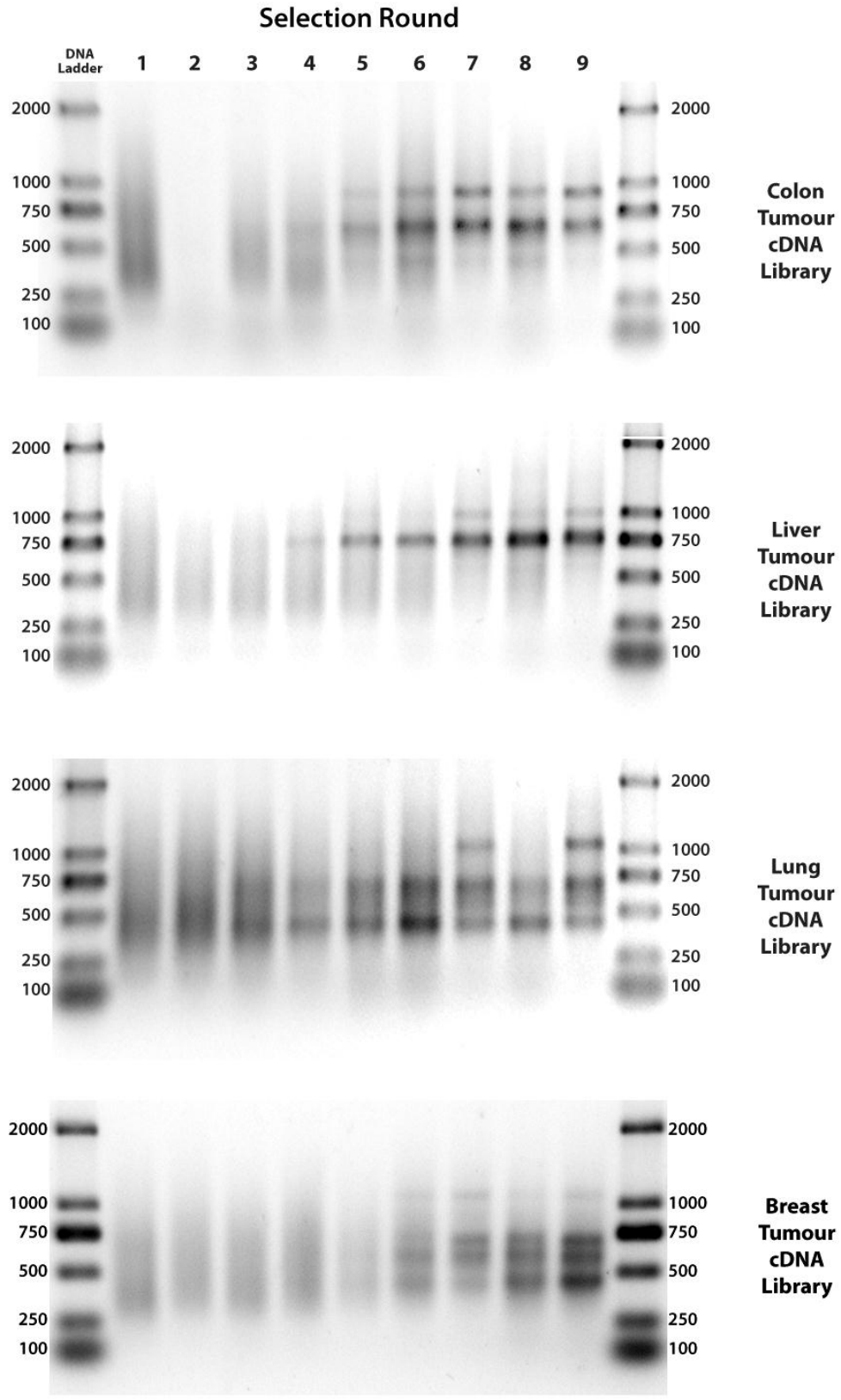


Figure 55: Agarose gel electrophoresis of phage DNA inserts amplified by PCR from colon tumour, liver tumour, lung tumour and breast tumour cDNA libraries after nine rounds of selection with biotinylated artesunate immobilsed on a neutravidin-coated microtitre plate.

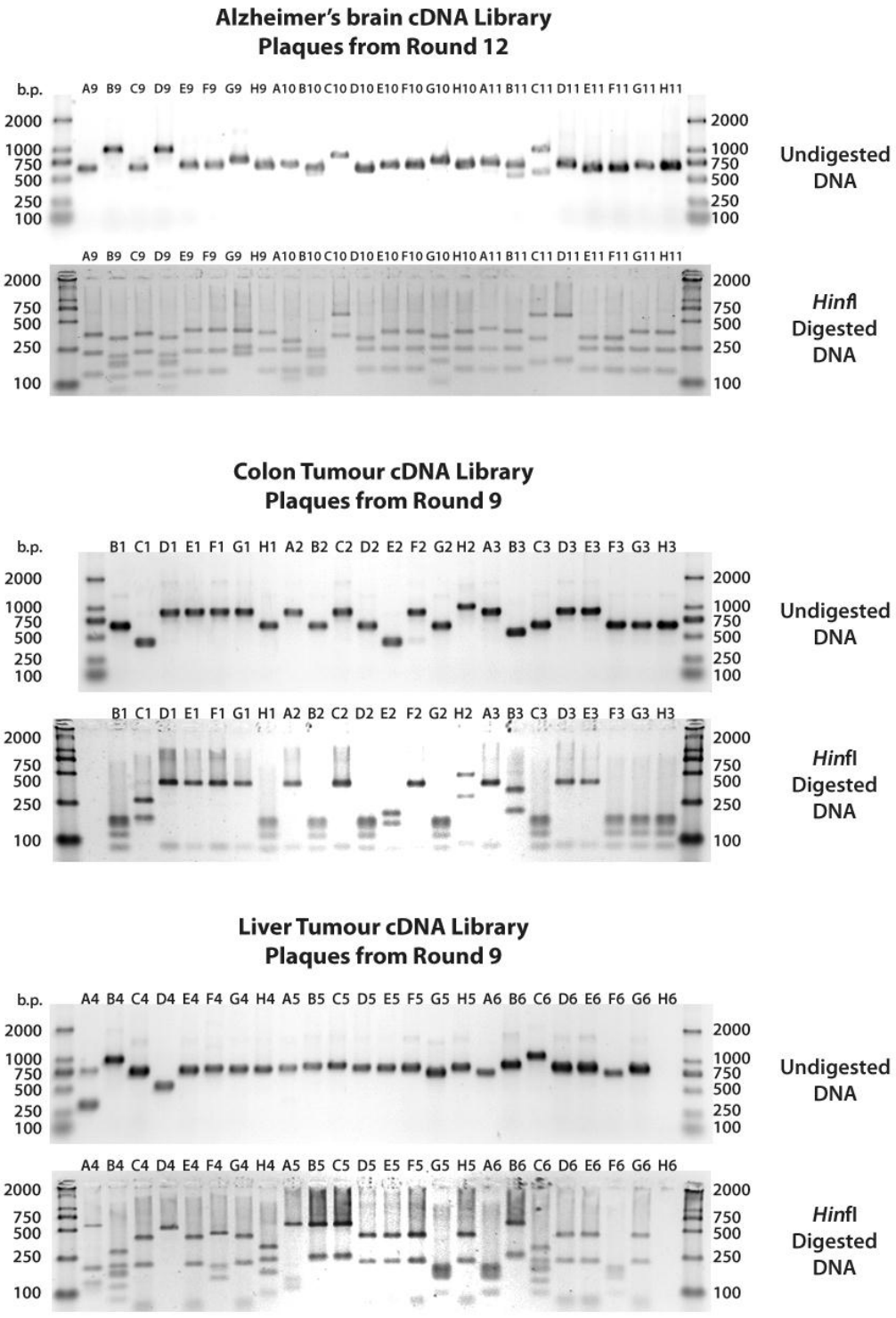


Figure 56: Agarose gel electrophoresis of PCR products obtained from Alzheimer's brain, colon tumour and liver tumour individual plaques after twelve or nine rounds of selection with biotinylated artesunate immobilised on a neutravidin-coated plate. The DNA inserts, which were amplified using generic T7 primers, were also digested with *Hinf*I to produce unique DNA fingerprints of each clone.

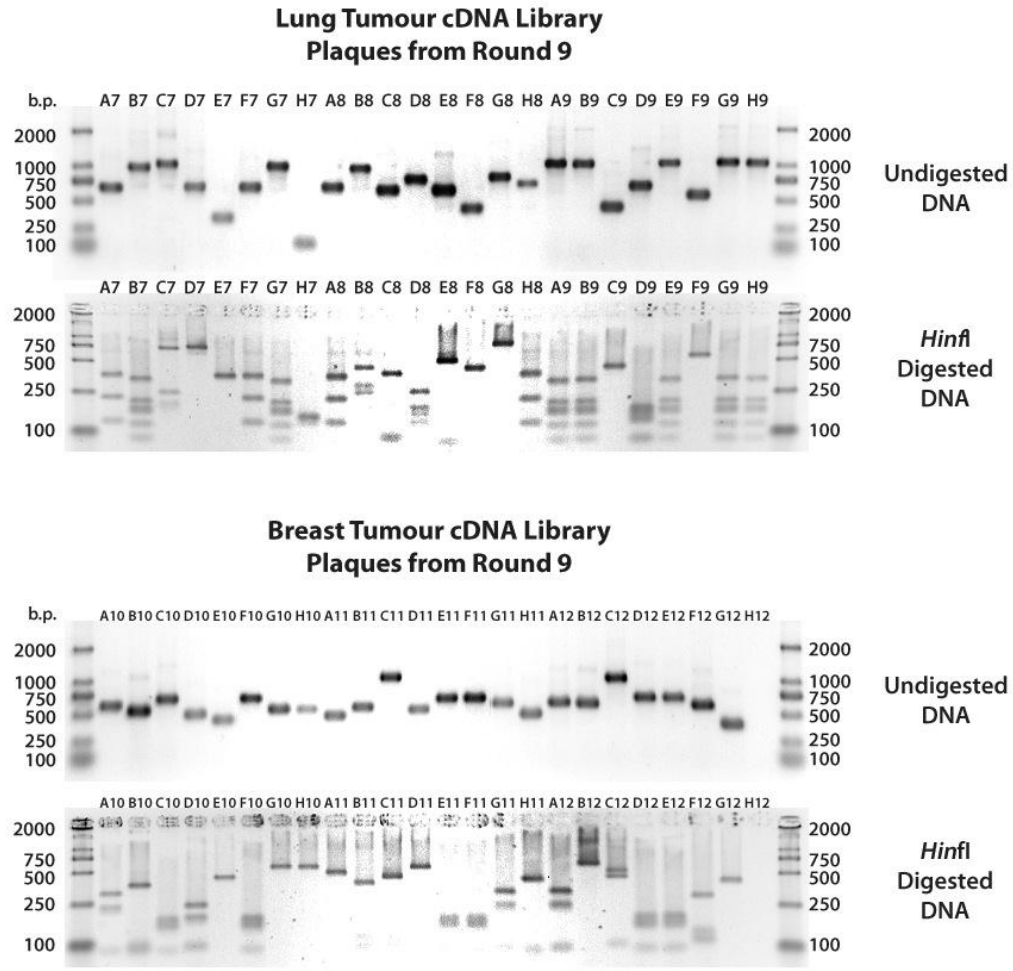


Figure 57: Agarose gel electrophoresis of PCR products obtained from lung and breast tumour individual plaques after nine rounds of selection with biotinylated artesunate immobilised on a neutravidin-coated plate. The DNA inserts, which were amplified using generic T7 primers, were also digested with *Hin*fI to produce unique DNA fingerprints of each clone.

Out of the six libraries tested, all five pathological derived were tending toward convergence, whereas the non-pathological colon library remained a mixture of many different clones, as indicated by a smear of PCR products during gel electrophoresis (see Figure 54). The colon tumour and liver tumour sublibraries from round 9 appeared to consist of two dominant clones each, whereas the selections with lung tumour and breast tumour libraries both gave rise to four major clones. In all cases the only clones that were in frame were shown to code for the same protein; BCL2-associated agonist of cell death (BAD), a protein essential in the apoptosis cascade, as a common target.

Table 23: DNA sequencing of PCR products obtained from individaul plaques after nine rounds of selection with biotin-artesunate immobilised on a neutravidin-coated PS microtitre plate.

Library	Plaques	Gene Identified from DNA Sequence	Notes	Frame
Col	-	*no convergence*	-	-
AB*	B9	**BCL2-associated agonist of cell death (BAD)**	entire CDS	1
AB*	A10, D10, E10	sphingomyelin phosphodiesterase 4	outside CDS	-
AB*	C10	melanoma antigen family F	last 210bp of 910bp	1
AB*	D11	chromosome 17, clone RP11-199F11, complete sequence	no CDS	-
CoT	B1	**BCL2-associated agonist of cell death (BAD)**	entire CDS	1
CoT	D1	complement component 1, q subcomponent, B chain	according protein after stop codon	2
CoT	E2	chromosome 8, clone RP11-550A5, complete sequence	no CDS	-
CoT	H2	sestrin 2 mRNA, complete cds	outside CDS	-
LiT	B4, C6	**BCL2-associated agonist of cell death (BAD)**	entire CDS	1
LiT	C4, G6	*no significant similarities*	-	-
LiT	B5	MRE11 meiotic recombination 11 homolog A (*S. cerevisiae*) (MRE11A)	contamination	-
LiT	C5	ankyrin repeat domain 49	gene in backwards	-
LuT	B7, D9, G9	**BCL2-associated agonist of cell death (BAD)**	entire CDS	1
LuT	D7	nuclear pore complex-interacting protein-like 3-like isoform 1	500bp of 3400bp CDS	1
LuT	F7	*no significant similarity*	-	-
LuT	D8	endothelial PAS domain protein 1, mRNA	gene in backwards	1
LuT	E8	cDNA DKFZp781B1548	only 10 aa displayed	-
LuT	G8	angiopoietin-like 4 (ANGPTL4) gene	800bp of 1.2kb CDS	2
BrT	C10	**BCL2-associated agonist of cell death (BAD)**	first 190bp of 510bp	1
BrT	C10	breast cancer-associated gene 1 protein (BCG1) mRNA, complete cds	370bp of 1.8kb CDS	1

* Plaques derived from this library were picked after 12 rounds of selection

Of the single plaques picked from each library, BAD represented by 2 out of 24 in Alzheimer's brain, 9 out 23 in colon tumour, 2 out of 23 in liver tumour, 8 out of 24 in

lung tumour and 6 out 23 in the breast tumour libraries. Although the phage clones incorporating parts of the BAD gene were never entirely dominant, the abundance of many BAD clones of different lengths in all five libraries is compelling.
Sequences of the BAD clones identified from different libraries were aligned to compare their start, end and length (see Figure 58).

```
                -15         -5              10         20         30
                ..|....| ....|....| ....|....| ....|....| ....|....|
AB_B9           -------- -SRDRAWAQS MFQIPEFEPS EQEDSSSAER GLGPSPAGDG
CoT_B1          -------- --SDRAWAQS MFQIPEFEPS EQEDSSSAER GLGPSPAGDG
LiT_C6          -------- ----GAWAQS MFQIPEFEPS EQEDSSSAER GLGPSPAGDG
LuT_B7          -------- -SRDRAWAQS MFQIPEFEPS EQEDSSSAER GLGPSPAGDG
LuT_G9          -------- ----GAWAQS MFQIPEFEPS EQEDSSSAER GLGPSPAGDG
BrT_C10         SEAAGPGQ GPRDRAWAQS MFQIPEFEPS EQEDSSSAER GLGPSPAGDG
BAD             -------- ---------- MFQIPEFEPS EQEDSSSAER GLGPSPAGDG
Consensus       -------- --rdrAWAQS MFQIPEFEPS EQEDSSSAER GLGPSPAGDG

                    40          50          60          70          80
               ....|....| ....|....| ....|....| ....|....| ....|....|
AB_B9          PSGSGKHHRQ APGLLWDASH QQEQPTSSSH HGGAGAVEIR SRHSSYPAGT
CoT_B1         PSGSGKHHRQ APGLLWDASH QQEQPTSSSH HGGAGAVEIR SRHSSYPAGT
LiT_C6         PSGSGKHHRQ APGLLWDASH QQEQPTSSSH HGGAGAVEIR SRHSSYPAGT
LuT_B7         PSGSGKHHRQ APGLLWDASH QQEQPTSSSH HGGAGAVEIR SRHSSYPAGT
LuT_G9         PSGSGKHHRQ APGLLWDASH QQEQPTSSSH HGGAGAVEIR SRHSSYPAGT
BrT_C10        PSGSGKHHSQ APGLLWDASH QQEQPTSSSH HGG------- ----------
BAD            PSGSGKHHRQ APGLLWDASH QQEQPTSSSH HGGAGAVEIR SRHSSYPAGT
Consensus      PSGSGKHHRQ APGLLWDASH QQEQPTSSSH HGGagaveir srhssypagt

                    90         100         110         120         130
               ....|....| ....|....| ....|....| ....|....| ....|....|
AB_B9          EDDEGMGEEP SPFRGRSRSA PPNLWAAQRY GRELRRMSDE FVDSFKKGLP
CoT_B1         EDDEGMGEEP SPFRGRSRSA PPNLWAAQRY GRELRRMSDE FVDSFKKGLP
LiT_C6         EDDEGMGEEP SPFRGRSRSA PPNLWAAQRY GRELRRMSDE FVDSFKKGLP
LuT_B7         EDDEGMGEEP SPFRGRSRSA PPNLWAAQRY GRELRRMSDE FVDSFKKGLP
LuT_G9         EDDEGMGEEP SPFRGRSRSA PPNLWAAQRY GRELRRMSDE FVDSFKKGLP
BrT_C10        ---------- ---------- ---------- ---------- ----------
BAD            EDDEGMGEEP SPFRGRSRSA PPNLWAAQRY GRELRRMSDE FVDSFKKGLP
Consensus      eddegmgeep spfrgrsrsa ppnlwaaqry grelrrmsde fvdsfkkglp

                   140         150         160
               ....|....| ....|....| ....|....| ....|...
AB_B9          RPKSAGIATQ MRQSSSWTRV FQSWWDRNLG RGSSAPSQ
CoT_B1         RPKSAGTATQ MRQSSSWTRV FQSWWDRNLG RGSSAPSQ
LiT_C6         RPKSAGTATQ MRQSSSWTRV FQSWWDRNLG RGSSAPSQ
LuT_B7         RPKSAGTATQ MRQSSSWTRV FQSWWDRNLG RGSSAPSQ
LuT_G9         RPKSAGTATQ MRQSSSWTRV FQSWWDRNLG RGSSAPSQ
BrT_C10        ---------- ---------- ---------- --------
BAD            RPKSAGTATQ MRQSSSWTRV FQSWWDRNLG RGSSAPSQ
Consensus      rpksagtatq mrqssswtrv fqswwdrnlg rgssapsq
```

Figure 58: Alignment of T7 phage-displayed BAD protein sequences isolated from Alzheimer's brain, colon tumour, liver tumour, lung tumour and breast tumour libraries with the human BAD analogue (EMBL bank entry: BC095431.1). The clone isolated from breast tumour resulted in a very short sequence and has not been included in the consensus alignment.

With the exception of the clone isolated from the breast tumour library (BrT_C10), all incorporated the entire CDS and all were in frame with the viral coat protein.

Cross-contamination is a possible source of the same protein being isolated from multiple libraries in the phage biopanning procedure (see Section 1.1.2.2.2). If this systemic error occurred, identical clones would appear throughout multiple libraries. The DNA and protein sequence differences of the isolated clones clearly indicate that we have successfully isolated multiple clones of the same protein from all the libraries. Clone AB_B9 varies from all other clones at position 137, where the naturally occurring threonine is replaced by an isoleucine. The responsible codon (ATA) from this clone differs from the corresponding codon (ACA) from other the clones (colon tumour, liver tumour, lung tumour, breast tumour). As this was not a mistake in sequencing, it is probably due to a single nucleotide polymorphism present in the donor from which the cDNA library was created. Also of interest, clone BrT_C10 only translates for only the first half of the BAD protein, suggesting the artesunate binding site is in the first half of the protein with an additional 18 amino acids at the 5'-end.

Detailed analysis of the amino acid composition of the consensus sequence of BrT_C10 protein and BAD (see Table 24) showed an even distribution of amino acids in most property categories.

Besides BAD, several other clones were present or dominant in the various libraries. From the AB library, the most dominant clone (e.g. AB_A10) is represented by 16 out of 24 plaques and incorporates the gene for sphingomyelin phosphodiesterase 4. However, the gene fragment inserted in the phage is outside the gene's protein coding sequence and results in a 21 aa short protein unrelated to the encoded protein. A protein-BLAST did not revealed any significant similarities involving more than 11 aa from the sequence.

The single plaques picked from the colon tumour library resulted in a second major clone apart from BAD: complement component 1, q subcomponent, B chain (C1QB) was present 9/23 times. However, DNA sequencing revealed that the gene's protein is only translated from reading frame 2 and furthermore the according coding sequence follows a hard stop codon (tga) present in the 5'-UTR.

Therefore, the CDS protein is not displayed on the phage surface. The protein resulting from reading frame 1 yields no significant similarity in a BLASTp search. The remaining four plaques are unique and were not sequenced.

Table 24: Amino acid composition of BAD and phage-displayed protein from the isolated plaque B9 from Alzheimer's brain library using biotin-artesunate immobilised on neutravidin-coated plates.

	BAD		BAD consensus with BrT_C10		
Residue	**Number**	**Mole%**	**Number**	**Mole%**	**Difference to BAD**
A = Ala	13	7.7	4.0	6.3	-6
D = Asp	8	4.8	3.0	4.8	-2
E = Glu	13	7.7	6.0	9.5	-3
F = Phe	6	3.6	2.0	3.2	0
G = Gly	19	11.3	9.0	14.3	-6
H = His	6	3.6	5.0	7.9	-1
I = Ile	2	1.2	1.0	1.6	0
K = Lys	4	2.4	1.0	1.6	0
L = Leu	7	4.2	3.0	4.8	-1
M = Met	4	2.4	1.0	1.6	0
N = Asn	2	1.2	0	0	0
P = Pro	15	8.9	7.0	11.1	-4
Q = Gln	11	6.5	6.0	9.5	-4
R = Arg	16	9.5	1.0	1.6	-3
S = Ser	27	16.1	12.0	19.0	-5
T = Thr	5	3.0	1.0	1.6	0
V = Val	3	1.8	0	0	0
W = Trp	5	3.0	1.0	1.6	-1
Y = Tyr	2	1.2	0	0	0

Property	**Residues**	**Number**	**Mole%**	**Number**	**Mole%**
Aliphatic	(A+I+L+V)	25	14.9	8	12.7
Aromatic	(F+H+W+Y)	19	11.3	8	12.7
Non-polar	(A+F+G+I+L+M+P+V+W+Y)	76	45.2	28	44.4
Polar	(D+E+H+K+N+Q+R+S+T)	92	54.8	35	55.6
Charged	(D+E+H+K+R)	47	28.0	16	25.4
Basic	(H+K+R)	26	15.5	7	11.1
Acidic	(D+E)	21	12.5	9	14.3

Except for two BAD clones, none of the other sequenced clones from the liver tumour (LiT) library resulted in any significant match. The most abundant clones (10 out of 23 plaques), as represented by C4 and G6, resulted in no significant similarities on the DNA level, nor did the resulting phage-displayed protein match any known protein.

The ankyrin repeat domain gene, which was found in clone C5, is inserted backwards into the phage DNA and therefore translates into a random peptide.
The sequencing results from the lung tumour screening gave rise to 8 clones with the BAD gene, encoding the entire CDS. Similar to the results from liver tumour library, a clone present multiple times turned out to show no significant similarities on the gene level, but when aligned with clone LiT-C4, also did not show any match to this clone. One clone (D7) perfectly matched the nuclear pore complex-interacting protein-like 3-like isoform 1, but only covers the last 15% (501 of 3368 bp) of the CDS. The DNA insert in the phage is translated and displayed by the phage in frame. The gene and protein entries on the NCBI database are predicted and no literature is available on this gene or the expressed protein.
BCL2-associated death promoter is ubiquitous in all human cells, and overexpressed in certain cancer types[384]. The fact that we did not isolate BAD from the colon library within nine rounds of selection could indicated extremely low protein abundance in the colon library compared to, for example, the colon tumour library or that the clone was lost in the early rounds.
Regarding the extra rounds of selection performed on the Alzheimer's brain library, the fingerprinting of round-12 single plaques revealed that a variety of clones were still present and the prolonged selection of three additional rounds did not result in one single clone or even one outnumbering others. This suggests that this library did not contain BAD or that it was lost in the early rounds of selection such that the library was unable to converge on a single clone.
The biopanning results of the artesunate probe yielded a number of clones but all appeared that carry parts of the BAD sequence and in each case translate for a phage-displayed protein in the correct reading frame.
This is a very intriguing finding, seeing that BAD as a member of the BCL-2 gene family is involved in initiating apoptosis - a common target for many anticancer drugs[385]. BAD is a member of the BCL-2 homology domain 3 (BH3)-only domain protein family and mostly lacks homology with the other family members such as BIM, BID, PUMA, HRK, NOXA or BIK besides the conserved domain. In the BAD protein, the BH3-domain sits at position 145-160 aa and is essential for its pro-apoptotic function[386,387]. If dephosphorylation of BAD occurs during the apoptosis

cascade, BAD hetero-dimerises with the anti-apoptotic BCL-2 or BCL-x_L to prevent either of those two factors from stopping the downstream apoptosis process[388].
The consensus protein sequence of all isolated phage clones expressing BAD (see Figure 58) do not contain the BH3-domain. This suggests that the binding site for artesunate is independent of the BH3-domain and this binding does not directly interfere with the formation of BAD-BCL-2/BCL-x_L hetero-dimers. However, at this point it remains unclear if and in which way the BAD function is impacted by artesunate.
An on-phage binding assay was performed to determine if the T7 phage-displayed BAD clones isolated had greater affinity for neutravidin-coated plates derivatised with biotinylated artesunate than for similar plates derivatised with a control compound (see Figure 59). A T7 wild type clone, lacking a cDNA insert, was used in parallel as a negative control.

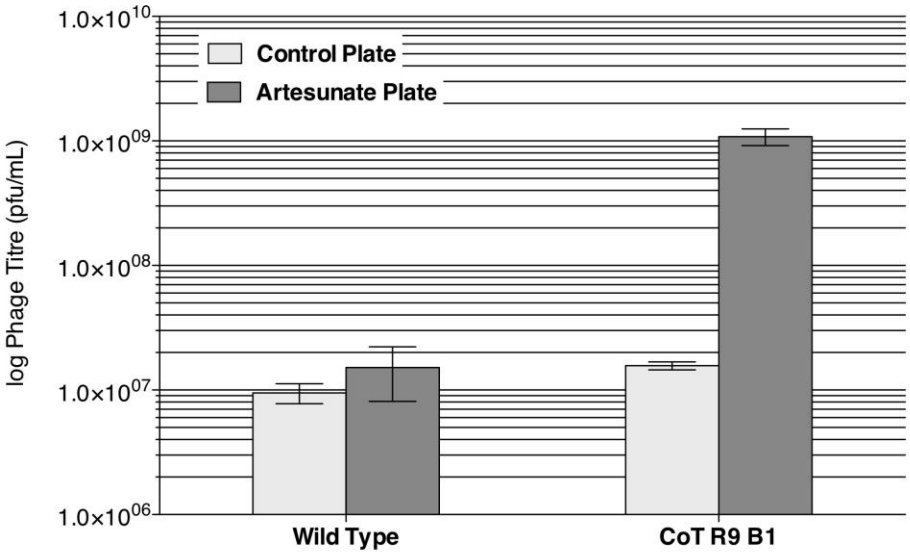


Figure 59: On-phage binding study comparing the affinity of the BAD-displaying phage clone B1 (from colon tumour sublibrary of round-9) for neutravidin-coated plates derivatised with a control compound and a similar plate derivatised with biotinylated artesunate.

Clearly, the BAD-displaying clone exhibited much higher affinity to the plate derivatised with biotin-artesunate than to the control plate. The phage titres observed from the artesunate-derivatised plates were 100-fold higher than those derived from the control probe-derivatised plates. The wild type clone displayed no significant affinity for either control or artesunate derivatised plates.
This finding strongly supports the conclusion that artesunate specifically binds to BAD and that this binding is responsible for the known anticancer activity of artesunate.

Chapter 5

Experimental

5.1 Materials and methods

Unless otherwise stated, all chemicals and reagents were received from Sigma-Aldrich (Castle Hill, Australia) and used without further purification. Chloroform, dichloromethane, diethyl ether, ethanol, ethyl acetate, light petroleum, methanol, tetrahydrofuran and toluene were obtained from ChemSupply (Australia), 1-butanol from Merck (Australia), and were glass distilled before use. HPLC grade acetonitrile was obtained from BDH/Merck (Germany), and was used without further purification.

5.1.1 Equipment

^{1}H NMR and ^{13}C NMR spectra were recorded in 5 mm Pyrex tubes (Wilmad, USA) on either a Bruker Avance DPX-400 400 MHz or DRX-600K 600 MHz spectrometer. All spectra were obtained at 25 °C, processed using Bruker Topspin 1.4 and referenced to residual solvent ($CDCl_3$ 7.26/77.16 ppm; methanol-d_4 3.31/ 49.00 ppm; DMSO-d_6 2.50/39.52 ppm). Infrared spectra were taken on a Perkin Elmer paragon 1000PC FTIR spectrometer, or Nicolet iS10 FT-IR Spectrometer (Thermo Scientific, Australia). UV measurements were performed on a NanoDrop 2000 UV-Vis-Spectrophotometer (Thermo Scientific, Australia). The specific optical rotation was measured on a Jasco P-1010 Polarimeter (Jasco, Japan).

Low resolution mass spectrometry was performed by electrospray ionisation (ESI) MS in positive polarity mode on a Shimadzu LC-20A prominence system coupled to a LCMS-2010 EV mass spectrometer using LCMSsolution 3.21 software. LC-MS experiments were carried out on a Gemini C18 column (Phenomenex, Australia) 150.0 × 2.00 mm, 110Å, 3 µm.

MS/MS experiments were performed on a HCT 3-D ion trap (Bruker Daltonics) running DataAnalysis 4.0 (Bruker). Full scan (*m/z* 200-2200) was followed by a data-dependent fragmentation of the two most abundant signals using alternating collision induced dissociation (CID) and electron transfer dissociation (ETD).

High resolution mass data were obtained from ESI in positive polarity mode on a Bruker 7T Fourier Transform Ion Cyclotron Resonance (FTICR) mass spectrometer, performed by the Mass Spectrometry Unit at University of Sydney, Australia.

High resolution-LC-MS experiments were carried out on a Thermo Orbitrap Elite mass spectrometer with Easy Spray source. The samples were loaded on a C18 trap column (Acclaim, Pepmap 100, 75 µm x 2cm, nanoviper, C18, 3 µm, 100 Å) and

switched online to the analytical column (Thermo, Easy Spray column, PepMap C18 column RSLC, C18, 2 µm, 100 Å, 50 µm x 15cm) after desalting. Samples were analysed in positive polarity mode using XCalibur 2.2 software. The HR-LC-ESI-MS experiment was performed by the Australian Proteome Analysis Facility, Sydney. HPLC analysis was performed on a Shimadzu 10AD-VP system running Class-VP 7.4 SP1 software. Analytical, semipreparative and preparative HPLC were performed on Gemini C18 HPLC columns (Phenomenex): Gemini-NX C18 250.0 × 4.6 mm, 110Å, 5 µm; Gemini C18 250.0 × 10.0 mm, 110Å, 10 µm; Gemini-NX C18 150.0 × 21.2 mm, 110Å, 5 µm.

TLC was performed with Merck Kieselgel 60 F_{254} plates with viewing under ultraviolet light (254 nm and 365 nm) and/or by heating after treatment with one of the following TLC staining solutions: 0.2 wt.% *p*-dimethylaminocinnam-aldehyde and 2% sulfuric acid in ethanol ("biotin stain"); 1.5% ninhydrin and 3.1% acetic acid in 1-butanol; 10% phosphomolybdic acid in ethanol; 1% potassium permanganate, 6.6% potassium carbonate, 0.84% sodium hydroxide in dist. water. Flash column chromatography was performed on silica gel (60 Å 0.040–0.063 mm, 230-400 mesh ASTM from Merck).

Disposable Strata-X Polymeric Reverse Phase cartridges were obtained from Phenomenex (Australia) and disposable C18 cartridges from Alltech (Grace, Australia) and pre-conditioned according to the manufacturer's recommendations. The freeze dryer system was from Labconco (USA).

5.2 Natural products

Sponges were collected by hand using SCUBA equipment or snorkel at Jervis Bay, NSW, Australia (S 35°08'07", E 150°43'33", at 1-4 m depth), in mid-May of each year of 2007-2009 under New South Wales Department of Primary Industries collecting permit No. P08/0006-1.1. The specimen ID indicates place and year of collection (e.g. JB07-S*x* indicates that the sponge sample *x* was collected at Jervis Bay in 2007). The samples were frozen upon collection and stored at -20 °C until further investigation. The sponges collected during this thesis project have only been subject to identification if they displayed high biological activity. Reference specimens were prepared according to Costantino[389] and are deposited at Department of Chemistry and Biomolecular Sciences, Macquarie University.

Ampicillin was obtained from Research Organics Inc. (USA), Carbenicillin from Fluka (Switzerland), and (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) from Alfa Aesar (USA). Bacto-Peptone and yeast extract powder were purchased from Difco Laboratories (USA), sodium chloride from Riedel-de Haën (Germany). Mueller Hinton broth (MHB) was from Bacto laboratories (Australia). Distilled water was used in the preparation of all aqueous extracts and partitions. Amberlite XAD-7 beads and Sephadex LH-20 were purchased from Fluka (Switzerland). 0.22 µm filters were used from Millex®-GP (UK). Filter paper discs (6 mm) used for disc diffusion assays were from Whatman (UK) and sterilised before use. Sterilised flat-bottomed 96 well plates were purchased from Greiner (Germany). The orbital shaker incubator was from Thermoline (Australia) and the blender (4 L) from Waring Commercial (USA).

5.2.1 Handling of sponge material and preparation for their identification

Digital photographs were taken from the collected material to facilitate identification of those species showing antibiotic or herbicidal activity. Additionally, spicule preparations were produced through bleach (sodium hypochlorite) and/or nitric acid digestion.

5.2.2 General extraction procedure for sponges

Once in the laboratory, all foreign material was removed from the sample before weighing, and extraction. Frozen (-20 °C) sponge samples were homogenised to a fine powder with a blender in distilled ethanol (500 mL). The supernatant was decanted

into a clean flask and the solid residue was blended with aqueous ethanol (80%; 500 mL). This process was repeated until the supernatant was no longer coloured or at least 3 times if the supernatant was transparent. The combined ethanol extracts were then filtered through a pad of celite, and the filtrate reduced to 300 mL *in vacuo*. The solution was partitioned against petroleum ether (3 × 300 mL). The combined petroleum ether washings were reduced to dryness *in vacuo*, redissolved in minimum amount of petroleum ether, transferred to pre-weighed glass vials and dried under nitrogen. The ethanol extracts were diluted with distilled water (500 mL). Brine (100 mL) was added to facilitate separation of the layers and the solution was partitioned against DCM (3 × 300 mL). The combined DCM extracts were reduced to dryness *in vacuo*, redissolved in minimum amount of DCM, transferred to pre-weighed glass vials and dried under nitrogen. Partitioning of the remaining extract against ethyl acetate (3 × 300 mL) was achieved by adding another portion of brine (100 mL). The combined ethyl acetate extracts were reduced to dryness *in vacuo*, redissolved in minimum amount of ethyl acetate, transferred to pre-weighed glass vials and dried under nitrogen. The aqueous ethanol was reduced in volume (Rotary evaporator) and partitioned against 1-butanol (3 × 300 mL). The combined butanol extracts were reduced to dryness *in vacuo*, transferred to pre-weighed glass vials and dried under nitrogen. To the remaining aqueous solution Amberlite XAD-7 beads were added and the solution was stirred gently for 6 h at room temperature. The beads were filtered from the solution, washed extensively with water and soaked in anhydrous ethanol (100 mL) for 18 h at 4 °C. Finally, the beads were filtered from the solution and the filtrate was reduced to dryness *in vacuo,* redissolved in minimum amount of ethanol, transferred to pre-weighed glass vials and dried under nitrogen.

5.2.3 Antibacterial Assays

Biosafety approval was obtained from the Macquarie University Biosafety Committee (approval number 5201000870).

5.2.3.1 Materials

The strains of bacteria used were *Staphylococcus aureus* ATCC 9144 (obtained from the CDS Reference Laboratory, Department of Microbiology, The Prince of Wales Hospital, NSW), *Escherichia coli* JM109 and *Pseudomonas aeruginosa* ATCC 27853 (both obtained from Professor Michael Gillings, Department of Biological Sciences, Macquarie University). Glycerol stock cultures (20% v/v glycerol) of these strains

Table 25: Preparation of media required for bioassays

Reagent	Ingredients	Instructions
LB Agar	10 g tryptone 5 g yeast extract 10 g NaCl 15 g agar Water to make 1 L of solution	- Adjust pH to 7.4 - Autoclave 20 min / 121 °C - Cool to 55 °C - Pour plates (~100) - Store plates at 4 °C
LB Agarose	10 g tryptone 5.2 g yeast extract 5.2 g NaCl 6 g agarose Water to make 1 L of solution	- Adjust pH to 7.4 - Autoclave 20 min / 121 °C - Cool to 55 °C - Store in 25 mL lots at 4 °C
MHB	2.0 g beef extract powder 17.5 g acid digest of casein 1.5 g soluble starch Water to make 1 L of solution	- Dissolve manufacturer's ready made powder in water - Autoclave 20 min / 121 °C
10 mM BES buffer	2.13 g BES Water to make 1 L of solution	- Adjust pH to 7.0 - Sterile filter (0.22 μm)
1/10 BYPN	2 g peptone 1 g yeast extract 1 g NaCl Water to make 100 mL of solution	- Adjust pH to 7.4 - Autoclave 20 min / 121 °C - Dilute with 900 mL BES buffer - Store in 50 mL lots at 4 °C
Hoagland's medium	0.246 g $MgSO_4.7H_2O$ 0.543 g $Ca(NO_3)_2.4H_2O$ 0.068 g KH_2PO_4 0.253 g KNO_3 0.5 mL Micronutrient solution Water to make 980 mL of solution **Micronutrient solution:** 2.86 g H_3BO_3 1.82 g $MnCl_2.4H_2O$ 0.22 g $ZnSO_4.7H_2O$ 0.09 g $Na_2MoO_4.2H_2O$ 0.09 g $CuSO_4.5H_2O$ Water to make 1 L of solution Fe•EDTA solution: 0.484 g $FeCl_3.6H_20$ 1.5 g EDTA Water to make 1 L of solution	- Adjust pH to 5.8 - Autoclave 20 min / 121 °C - Add 20.0 mL Fe•EDTA solution aseptically

were kept at -80 °C. Initial cultures were prepared by streaking a small quantity of the frozen stock onto an LB agar plate and incubation at 37 °C for 16 h. The plates were stored at 4 °C and used to inoculate subsequent cultures for up to one month. Overnight cultures were generally prepared by inoculating the according medium with a single bacterial colony from the agar plates, and subsequent incubation at 37 °C for 16 h with shaking (100-150 rpm). The optical density at 600nm (OD_{600}) was adjusted to 0.08-0.1 to produce an inoculum density of 1.0×10^8 cfu/mL (OD).

5.2.3.2 Disc diffusion

For the disc diffusion assay, the saturated overnight cultures were prepared in Luria Broth (LB) (15 mL). Sterile paper discs (∅ 6 mm) were impregnated with the sponge extracts (2 × 25 µL; 10 mg/mL 1:1 acetone:water), antibiotic (10 µL; 1 mg/mL) or solvent control (2 × 25 µL; 1:1 acetone:water) and dried hereafter. LB agarose medium (25 mL) was inoculated with the above inoculum (1 mL) at 45-50 °C, vortexed and poured (5 mL) onto fresh LB agar plates to give a thin layer (2 mm). The impregnated paper discs were pressed into the solidifying LB agarose layer. The plates were incubated at 37 °C for 16 h and the diameter of the zone of inhibition was measured with a ruler. Measurements were rounded off to the next mm.

5.2.3.3 Turbidity

For the MTT microdilution assay, all bacterial strains were grown overnight in MHB (15 mL). The assay was performed in sterile, clear flat-bottom-96-well microtitre plates. The arrangement of samples and controls was outlined according to Appendino *et al*.[184]. The sponge extracts (10.0 mg), pure HPLC-purified bromotyrosines (1.0 mg/mL) or the antibiotics (1 µg) were dissolved in DMSO (200 µL) and the final volume was made up to 1 mL with distilled water. Using a 96 well microtitre plate, MHB was dispensed into wells 1-11 (125 µL each) for each row, the extract or antibiotic solution was added to well 1 (125 µL; in different rows for each extract) and mixed thoroughly, after which 125 µL was taken out and dispensed to the next well (*i.e.* well 2). This process of two-fold serial dilution was carried out until well 10, and skipping well 11, the final volume was dispensed into well 12. Again, 125 µL each of the bacterial inoculum was dispensed into wells 1 to 11 leaving well 12 free of inoculum. Well 11 was free of the test compound or the antibiotic, thus this acted as a positive growth control. Similarly, well 12 served as sterile control of the assay. 5% DMSO was also included as a negative control. The optical density at 600

nm (OD_{600}) was measured and the plate was incubated at 37 °C for 18 hours. After incubation, the OD_{600} was determined again and the pre-incubation values subtracted. The resulting optical densities were plotted against molar concentration of the compounds and analysed in GraphPad Prism to yield EC_{50} values.

5.2.3.4 MTT microdilution assay

To the plates resulting from 5.2.3.3 (after incubation), 20 µL of a solution of MTT (5 mg/mL methanol) was added to each well and again incubated for 30 minutes. The MIC was determined as the lowest concentration of the test compound or antibiotic that showed no visible colour change from yellow to blue/purple.

5.2.4 Herbicidal

5.2.4.1 Maintenance of Wolffia arrhiza

Duckweed (*Wolffia arrhiza*) was cultured and maintained in sterile Hoagland's medium[390]. All duckweed used for this study originated from a single *W. arrhiza* (an aggregate of one mother and one daughter) frond. The plants were grown in sterile Hoagland's medium (50 mL, see Table 25) in sterile Petri dishes in a growth cabinet. The sterile stock was incubated for 14 days in the growth cabinet under controlled lighting and temperature conditions of 16:8 light:dark cycle at 25 °C and 278 µmol/sec lumen, after which dish and media were refreshed.

The growth of sterile cultures of duckweed required closed containers such as Petri dishes fitted with a low evaporation lid, transfer of individual plants was conducted with a flamed bacterial loop as well as a cool light sources or in a growth incubators with temperature control. In case of contamination or infection by fungi individual plants were submerged (10 sec) in 10-50% sodium hypochlorite, immediately washed (submerged in sterile media) for 30 sec and each plant placed into individual sterile growth containers.

5.2.4.2 Herbicidal assay

Screening assays were conducted in sterile transparent 96 well microtitre plates with a standard flat bottom. Each well contained Hoagland's medium (175 µL) and the natural product extract dissolved in a solution of acetone:water (20 µL; 1:1) or DMSO:water (20 µL; 1:9). The control wells either contained acetone:water (20 µL; 1:1) or DMSO:water (20 µL; 1:9), the positive control contained glyphosate (*N*-(phosphonomethyl)glycine; 2 µL of 458 mg/mL water) the negative control 20 µL of water. Serial dilutions were made for all compounds down to 10^{-5} with respect to the

original concentration (10 mg/mL). Each well was inoculated with two-fronds of duckweed of the same age and approximately the same size. The 96-well microtitre plates were incubated for 9 days in a growth cabinet at controlled lighting and temperature conditions of 16:8 light:dark cycle at 25 °C and 278 μmol/sec lumen. All samples were tested as duplicates.

5.2.5 Isolation of bioactive marine natural products

5.2.5.1 Pseudoceratina purpurea

The colour of the sponge changed from yellow (underwater) to a deep purple after exposed to air for some time. The individual was frozen and stored at -20 °C until extraction. All foreign material was removed from the sponge. The frozen sponge (654 g) was homogenised to a fine powder with a blender in distilled ethanol (2000 mL). The supernatant was decanted into a clean flask and the solid residue was blended with aqueous ethanol (80%; 1000 mL). This process was repeated until the supernatant was no longer coloured (4 ×). The combined ethanol extracts were then filtered through a pad of celite, and the ethanol removed *in vacuo* to yield approximately 900 mL aqueous extract. An aliquot (20 mL) was put aside, freezedried and stored at -80 °C.

The remaining solution was partitioned against petroleum ether (3 × 500 mL). The combined petroleum ether washings were reduced to dryness *in vacuo*, redissolved in minimum amount of petroleum ether, transferred to pre-weighed glass vials and dried under nitrogen to yield a dark green-brown residue (0.9076 g). Brine (150 mL) was added to facilitate separation of the layers and the solution partitioned against ethyl acetate (5 × 500 mL). The combined EtOAc extracts were reduced to dryness *in vacuo*, redissolved in minimum amount of EtOAc, transferred to pre-weighed glass vials and dried under nitrogen to yield a brown-purple residue (6.5420 g). The aqueous ethanol was reduced in volume again (~ 600 mL; rotary evaporator, 40 °C) and partitioned against 1-butanol (3 × 200 mL). The combined butanol extracts were reduced to dryness *in vacuo*, transferred to pre-weighed glass vials and dried under nitrogen yield a light brown residue (0.2241 g). To the remaining aqueous solution Amberlite XAD-7 beads were added and the solution was stirred gently for 6 h at room temperature. The beads were filtered from the solution, washed extensively with water and soaked in anhydrous ethanol (100 mL) for 18 h at 4 °C. Finally, the beads were filtered from the solution and the filtrate was reduced to dryness *in*

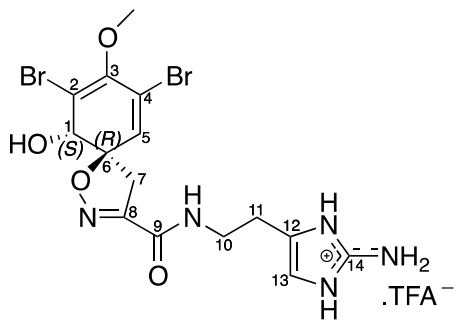
vacuo, redissolved in minimum amount of ethanol, transferred to pre-weighed glass vials and dried under nitrogen to yield a light yellow residue (5.7681 g). All extracts were stored at -80 °C until further analysis.

The crude ethyl acetate extract (3.5 g) was subjected to gel permeation chromatography on Sephadex LH-20 (1:1 chloroform:methanol) and 90 fractions were collected (see Section 8.1.1.1.1).
Fractions 9-22 were combined (based on TLC) and reduced to dryness *in vacuo* to give a dark sticky residue (1.66 g). The resulting dark residue was prepared for analysis by HPLC(-MS). Therefore an equivalent (100 mg) was loaded in portions onto a disposable Strata-X cartridge (500 mg bed weight), washed with distilled water (3 mL) and eluted with 80% acetonitrile/water (10 mL). The solvent was evaporated *in vacuo* to yield a brown residue (74 mg), which was redissolved in 5% acetonitrile:water (5 mL). An aliquot (50 µL) was diluted with 10% acetonitrile:aqueous FA (0.1%) and analysed via LR-LC-MS (Gemini C18 column, 0.2 mL/min, gradient from 10-95% acetonitrile:aqueous FA (0.1%) over 35 min). The remaining solution was separated via HPLC (preparative C18 column, 9.99 mL/min, gradient from 18-40% acetonitrile:water over 40 min). For HR-LC-MS analysis 2 µL samples (20 µg/mL) were eluted over 30 min from 5% B to 90% B (Buffer A: 0.1% formic acid, 99.9% milliQ water, Buffer B: 0.1% formic acid, 99.9% acetonitrile) at a flow rate of 300 nL/min.

5.2.5.2 Brominated tyrosines

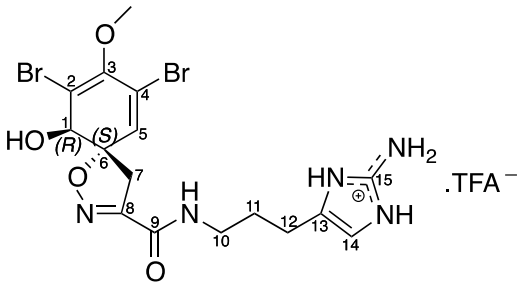
For all known, previously published compounds, the atoms were numbered according to the original publications for NMR assignments. For NMR data (HSQC, HMBC, COSY) of new bromotyrosine compounds (**55** and **60**) see Sections 8.1.3 and 8.1.4.

5.2.5.2.1 (–)-Pseudoceratinin A (54)[200]



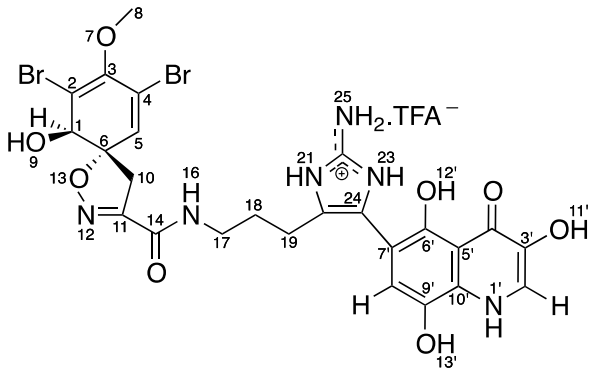
Unstable light yellow solid (15.4 mg; 0.038% of wet weight). $[\alpha]_D^{20}$ –154° (*c* 0.5, MeOH) Lit.[200] –158°. ^{1}H-NMR (600 MHz, DMSO-d_6) δ 12.10 (s, NH-13), 11.80 (s, NH-12), 8.60 (t, *J* = 8.6 Hz, NH-9), 7.40 (NH_2-14), 6.59 (s, H-13), 6.55 (bs, C1-OH), 6.54 (s, H-5), 3.89 (s, H-1), 3.62 (s, OMe), 3.59 (d, *J* = 18.2 Hz, H7a), 3.35 (q, *J* = 6.4 Hz, (H-10)$_2$), 3.17 (d, *J* = 18.1 Hz, H-7b), 2.59 (t, *J* = 6.5 Hz, (H-11)$_2$). ^{13}C-NMR (150 MHz, DMSO-d_6) δ 159.5 (C-9), 154.9 (C-8), 147.6 (C-3), 147.5 (C-14), 131.7 (C-5), 124.7 (C-12), 121.3 (C-2), 113.6 (C-4), 109.9 (C-13), 90.7 (C-6), 74.1 (C-1), 60.1 (MeO), 39.5 (C-7), 37.8 (C-10), 24.8 (C-11). Mass spectrum (ESI+) *m/z*: isotopic cluster 490:492:494 (in ratio 1:2:1).

5.2.5.2.2 (+)-Aerophobin-2 (44)[201]



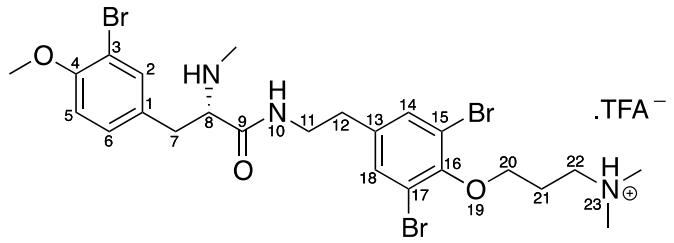
Off-white solid (9.1 mg; 0.022% of wet weight). $[\alpha]_D^{20}$ +128° (*c* 0.55, MeOH), Lit.[201] +139°. ^{1}H-NMR (600 MHz, DMSO-d_6) δ 12.11 (s, NH-14), 11.70 (s, NH-13), 8.57 (t, *J* = 5.9 Hz, NH-9), 7.37 (s, NH_2-15), 6.57 (s, H-14), 6.56 (s, H-5), 3.90 (s, H-1), 3.63 (s, OMe), 3.60 (d, *J* = 17.8 Hz, H-7a), 3.18 (d, *J* = 17.8 Hz, H-7b), 3.16 (q, *J* = 6.6 Hz, (H-10)$_2$), 2.39 (t, *J* = 7.4 Hz, (H-12)$_2$), 1.69 (m, (H-11)$_2$). ^{13}C-NMR (150 MHz, DMSO-d_6) δ 159.5 (C-9), 155.0 (C-8), 147.6 (C-3), 147.3 (C-15), 131.7 (C-5), 126.8 (C-13), 121.3 (C-2), 113.6 (C-4), 109.3 (C-14), 90.7 (C-6), 74.1 (C-1), 60.1 (MeO), 39.8 (C-7), 38.5 (C-10), 27.8 (C-11), 22.0 (C-12). Mass spectrum (ESI+) m/z: isotopic cluster 504:506:508 (in ratio 1:2:1).

5.2.5.2.3 (+)-Ceratinadin D (55)



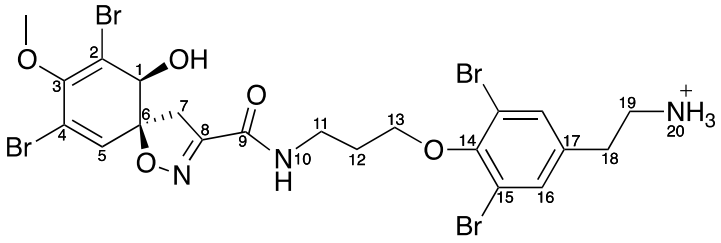
Unstable brown amorphous solid (4.5 mg; 0.011% of wet weight). $[\alpha]_D^{20}$+52° (*c* 0.45, MeOH), Lit.[217] +51° [(+)-ceratinadin A]. λ_{max} (MeOH) 234, 276, 337 nm. ν_{max} (Neat film) 3420 (m), 3380 (m), 3290 (m), 1677 (s), 1432 (s), 1202 (d), 1135 (s), 1047 (s), 1025 (br), 802 (s), 764 (s), 722 (s) cm^{-1}. ^{1}H-NMR (600 MHz, DMSO-d_6) δ 14.8 (bs, H-12'), 12.15 (bs, H-23), 11.93 (bs, H-21), 11.76 (d, *J* = 6.4 Hz, H-1'), 10.25 (bs, H-13'), 8.93 (bs, H-11'), 8.51 (t, *J* = 5.8 Hz, H-16), 7.27 (bs, H-25), 7.64 (d, *J* = 6.4 Hz, H-2'), 6.83 (s, H-8'), 6.54 (s, H-5), 6.36 (d, *J* = 7.9 Hz, H-9), 3.90 (d, *J* = 7.6 Hz, H-1), 3.63 (s, OMe), 3.59 (d, *J* = 3.6 Hz, H-10a), 3.15 (d, *J* = 3.6 Hz, H-10b), 3.13 (m, H-17), 2.51 (s, H-19), 1.73 (m, H-18). ^{13}C-NMR (150 MHz, DMSO-d_6) δ 173.2 (C-4'), 158.9 (C-14), 154.5 (C-11), 149.6 (C-6'), 147.2 (C-3), 139.9 (C-3'), 137.4 (C-9'), 131.2 (C-5), 128.9 (C-10'), 124.2 (C-2'), 122.4 (C-20), 120.9 (C-2), 117.8 (C-24), 113.7 (C-8'), 113.1 (C-4), 112.6 (C-5'), 90.1 (C-6), 73.6 (C-1), 59.7 (C-8), 39.0 (C-10), 38.2 (C-17), 28.1 (C-18), 21.6 (C-19). Mass spectrum (ESI+) *m/z*: isotopic cluster 695:697:699 (in ratio 1:2:1). (HRESI+) Found *m/z*: 695.0067, $C_{25}H_{25}N_6O_8{}^{79}Br_2$ requires 695.0101.

5.2.5.2.4 (+)-Suberedamine B (56)[202]



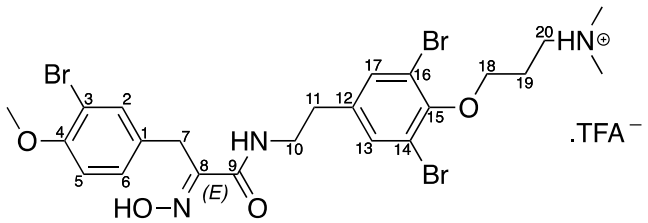
Slighty yellow solid (2.0 mg; 0.005% of wet weight). $[\alpha]_D^{20}$ +19° (*c* 0.1, MeOH), Lit.[202] +16°. ^{1}H-NMR (400 MHz, methanol-d_4) δ 7.45 (d, *J* = 2.0 Hz, H-2), 7.42 (s, H-14 and H-18), 7.17 (dd, *J* = 8.4 Hz, 2.2 Hz, H-6), 6.90 (d, *J* = 8.4, H-5), 4.01 (t, *J* = 6.1 Hz, H-8), 3.84 (s, OMe), 3.81 (t, *J* = 5.6 Hz, $(H\text{-}20)_2$), 3.44 (t, *J* = 7.2 Hz, $(H\text{-}22)_2$), 2.74 (m, $(H\text{-}7)_2$ and $(H\text{-}12)_2$), 2.38 (s, NMe_2), 2.07 (m, $(H\text{-}21)_2$). ^{13}C-NMR (100 MHz, methanol-d_4) δ 165.8 (C-9), 155.8 (C-4), 153.0 (C-16), 139.8 (C-13), 134.8 (C-2), 134.3 (C-14 and C-18), 131.7 (C-6), 130.3 (C-1), 118.9 (C-15 and C-17), 113.2 (C-5), 112.2 (C-3), 72.5 (C-20), 57.5 (C-22), 56.7 (OCH_3), 56.6 (C-8), 42.7 (C-8-N<u>Me</u>), 45.2 (NMe_2), 41.3 (C-7), 35.1 (C-12), 33.3 (C-11), 28.7 (C-21). Mass spectrum (ESI+) *m/z*: isotopic cluster 648:650:652:654 (in ratio 1:3:3:1).

5.2.5.2.5 (–)-Hexadellin A (57)[203]



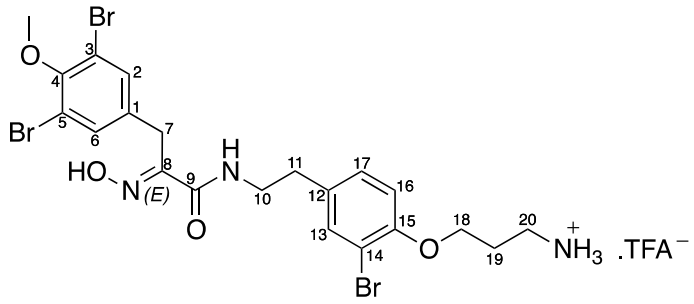
Slight-pink solid (8.1 mg; 0.020% of wet weight). $[\alpha]_D^{20}$ –19° (*c* 0.5, MeOH). ^{1}H-NMR (400 MHz, DMSO-d_6) δ 8.55 (t, *J* = 5.6 Hz, H-10), 7.82 (bs, (H-20)$_2$), 7.57 (bs, (H-16)$_2$), 6.56 (bs, H-5), 6.36 (bs, C1-OH), 3.94 (t, *J* = 6.2 Hz, (H-13)$_2$), 3.90 (s, H-1), 3.62 (s, OMe), 3.58 (d, *J* = 18.2 Hz, H-7a), 3.19 (d, *J* = 18.2 Hz, H-7), 3.38 (m, H-11), 3.05 (m, (H-19)$_2$), 2.81 (t, *J* = 7.2 Hz, (H-18)$_2$), 1.98 (m, (H-12)$_2$). ^{13}C-NMR (100 MHz, DMSO-d_6) δ159.4 (C-9), 155.0 (C-8), 151.7 (C-14), 147.6 (C-3), 137.3 (C-17), 133.7 (C-16), 131.7 (C-5), 121.3 (C-2), 118.1 (C-15), 113.6 (C-4), 90.7 (C-6), 74.1 (C-1), 71.7 (C-13), 60.1 (OMe), 39.0 (C-7), 39.0 (C-19), 36.7 (C-11), 32.0 (C-18), 29.9 (C-12). Mass spectrum (ESI+) *m/z*: isotopic cluster 714:716:718:720:722 (in ratio 1:4:6:4:1).

5.2.5.2.6 Aplysamine 2 (58)[204]



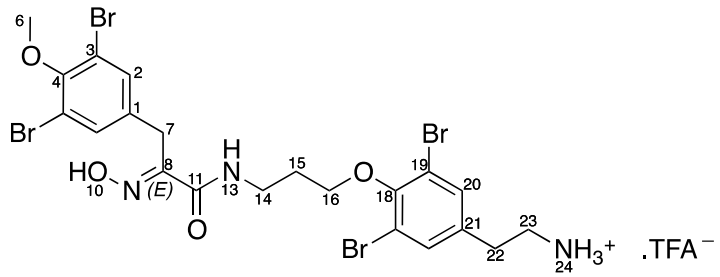
Light brown solid (9.0 mg; 0.022% of wet weight). ^{1}H-NMR (400 MHz, DMSO-d_6) δ 11.90 (s, N-OH), 8.05 (t, *J* = 6.0 Hz, NH-9), 7.48 (s, H-13 and H-17), 7.37 (d, *J* = 2.0 Hz, H-2), 7.11 (dd, *J* = 8.5 Hz, 2.1 Hz, H-6), 6.97 (d, *J* = 8.5 Hz, H-5), 3.96 (t, *J* = 5.9 Hz, (H-18)$_2$), 3.78 (s, OMe), 3.70 (s, (H-7)$_2$), 3.35 (m, (H-10)$_2$), 3.34 (m, (H-20)$_2$), 2.83 (s, (NMe$_2$), 2.72 (t, *J* = 7.0 Hz, H-11), 2.18 (m, (H-19)$_2$). ^{13}C-NMR (100 MHz, DMSO-d_6) δ 110.1 (C-1), 133.0 (C-2), 112.6 (C-3), 153.0 (C-4), 113.8 (C-5), 129.2 (C-6), 27.8 (C-7), 151.8 (C-8), 163.3 (C-9), 39.6 (C-10), 33.4 (C-11), 139.4 (C-12), 133.1 (C-13 and C-17), 117.2 (C-14 and C-16), 150.2 (C-15), 70.2 (C-18), 24.8 (C-19), 39.6 (C-20), 56.2 (OMe), 42.4 (NMe$_2$). Mass spectrum (ESI+) *m/z*: isotopic cluster 648:650:652:654 (in ratio 1:3:3:1).

5.2.5.2.7 16-Debromoaplysamine 4 (59)[205]



Light brown solid (1.6 mg; 0.004% of wet weight). ^{1}H-NMR (400 MHz, DMSO-d_6) δ 11.98 (s, N-OH), 8.16 (t, J = 5.9 Hz, NH-9), 7.77 (s, NH), 7.47 (s, H-13), 7.45 (s, H-1 and H-5), 7.19 (dd, J = 8.4 Hz, 2.1 Hz, H-17), 7.00 (d, J = 8.4 Hz, H-16), 3.99 (t, J = 6.1 Hz, $(H\text{-}18)_2$), 3.75 (s, $(H\text{-}7)_2$), 3.74 (s, (OMe), 3.34 (m, $(H\text{-}20)_2$), 3.00 (t, J = 7.8 Hz, $(H\text{-}10)_2$), 2.77 (t, J = 7.7 Hz, $(H\text{-}11)_2$), 1.92 (m, $(H\text{-}19)_2$). ^{13}C-NMR (100 MHz, DMSO-d_6) δ 163.3 (C-9), 153.6 (C-15), 151.8 (C-3), 151.1 (C-8), 136.3 (C-6), 133.0 (C-13), 132.8 (C-1, C-5), 131.2 (C-12), 129.2 (C-17), 117.0 (C-2, C-4), 113.7 (C-16), 111.3 (C-14), 66.7 (C-18), 60.2 (OMe), 39.8 (C-10), 36.1 (C-20), 31.7 (C-11), 28.7 (C-19), 27.9 (C-7). Mass spectrum (ESI+) *m/z*: isotopic cluster 620:622:624:626 (in ratio 1:3:3:1).

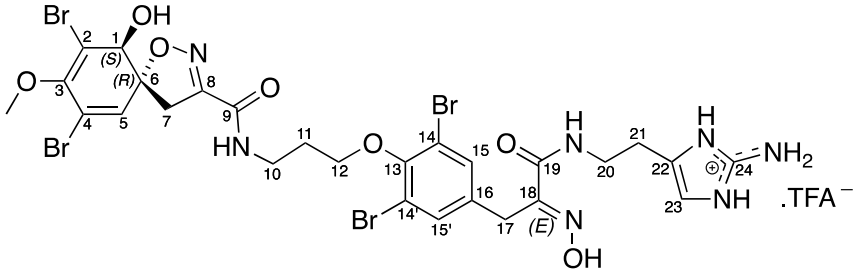
5.2.5.2.8 New Bromotyrosine (60)



Light yellow solid (0.9 mg; 0.002% of wet weight). λ_{max} (MeOH) 209, 278 nm. ν_{max} (Neat film) 3572 (m), 3187 (m), 3061 (m), 2930 (m), 1721 (m), 1685 (s) 1672 (m), 1642 (m), 1631 (m), 1546 (m), 1422 (s), 1258 (s), 1201 (br), 1001 (s), 835 (s), 734 (s), 721 (s) cm^{-1}. ^{1}H-NMR (400 MHz, DMSO-d_6) δ 12.02 (s, H-10), 8.12 (t, J = 6.0 Hz, H-13), 7.80 (bs, $(H\text{-}24)_2$), 7.55 (s, $(H\text{-}20)_2$), 7.44 (s, $(H\text{-}2)_2$), 3.88 (t, J = 6.4 Hz, $(H\text{-}16)_2$), 3.76 (s, $(H\text{-}9)_2$), 3.75 (s, OMe), 3.38 (m, $(H\text{-}14)_2$), 3.05 (m, $(H\text{-}23)_2$), 2.81 (t, J = 7.4 Hz, $(H\text{-}22)_2$), 1.96 (m, $(H\text{-}15)_2$). ^{13}C-NMR (100 MHz, DMSO-d_6) δ 163.0 (C-11), 151.8 (C-4), 151.3 (C-18), 151.0 (C-8), 136.8 (C-21), 136.3 (C-1), 133.2 ($(C\text{-}20)_2$), 132.9 ($(C\text{-}2)_2$), 117.6 ($(C\text{-}19)_2$), 117.1 ($(C\text{-}3)_2$), 71.3 (C-16), 60.4 (OMe), 27.9 (C-7), 39.4 (C-23), 36.2 (C-14), 31.5 (C-22), 29.6 (C-15). Mass spectrum (ESI+) *m/z*: isotopic cluster

698:700:702:704:706 (in ratio 1:4:6:4:1). (HRESI+) Found 697.8527, $C_{21}H_{24}{}^{79}Br_4N_3O_4$ requires 697.8500.

5.2.5.2.9 (–)-Purealin (61)[206]



Light brown solid (1.9 mg; 0.005% of wet weight). $[\alpha]_D^{20}$ –82° (*c* 0.1, MeOH), Lit.[206] –85°. ^{1}H-NMR (400 MHz, DMSO-d_6) δ 12.03 (s, N-OH), 11.85 (s, NH-22), 8.57 (bs, NH-9), 8.15 (bs, NH-19), 7.44 (s, H-15 and H-15'), 7.35 (bs, NH_2-24), 6.57 (d, *J* = 10.6 Hz, H-5), 6.56 (s, H-23), 6.38 (bs, C1-OH), 3.90 (d, *J* = 8.1 Hz, H-1), 3.75 (s, $(H\text{-}17)_2$), 3.63 (s, (OMe), 3.61 (d, *J* = 18.1 Hz, H-7a), 3.49 (m, $(H\text{-}12)_2$), 3.38 (m, $(H\text{-}10)_2$), 3.36 (m, $(H\text{-}20)_2$), 3.20 (d, *J* = 18.3 Hz, H-7b), 2.59 (m, $(H\text{-}21)_2$), 1.97 (m, $(H\text{-}11)_2$). ^{13}C-NMR (100 MHz, DMSO-d_6) δ 163.3 (C-19), 158.9 (C-9), 154.5 (C-8), 151.7 (C-13), 151.0 (C-18), 147.1 (C-3), 136.3 (C-16), 132.9 (C-15 and C-15'), 131.3 (C-5), 124.4 (C-22), 120.8 (C-2), 117.1 (C-14 and C-14'), 113.1 (C-4), 109.2 (C23), 90.2 (C-6), 73.6 (C-1), 69.8 (C-12), 59.4 (OMe), 39.4 (C-7), 37.3 (C-20), 36.2 (C-10), 29.4 (C-11), 27.9 (C-17), 24.5 (C-21). Mass spectrum (ESI+) *m/z*: isotopic cluster 880:882:884:886:888 (in ratio 1:4:6:4:1).

5.2.5.3 Tylodina corticalis

The opisthobranch *Tylodina corticalis* was found feeding on the sponge *Pseudoceratina purpurea*. The individual was separated from the sponge, frozen at -20 °C and once returned to the lab stored at -80 °C until extraction. Upon collection and freezing, the opisthobranch secreted a white-yellow slimy liquid, which was extracted along with the animal by soaking overnight in ethanol (5 × 50 mL) at -20 °C. All extracts were combined, evaporated to yield a brown oil (327.6 mg). An aliquot (27.6 mg) was put aside and kept at -80 °C. 300 mg were redissolved in water (10 mL) and a minimum amount of ethanol (1.5 mL). The insoluble remainders dissolved upon partitioning with petroleum ether (3 × 30 mL). The combined petroleum ether washings were reduced to dryness *in vacuo*, to yield a brown solid (40.1 mg). From the aqueous ethanol extracts most of the ethanol was evaporated, diluted with

distilled water (10 mL) and partitioned against ethyl acetate (4 × 30 mL). The combined ethyl acetate extracts were reduced to dryness *in vacuo*, yielding a light brown solid (12.0 mg). The aqueous ethanol was partitioned against 1-butanol (5 × 30 mL). The combined butanol extracts (95.6 mg) were reduced to dryness *in vacuo* which resulted in the yellow solution turning dark-purple, suggesting the presence of the verongid sponge pigment uranidine[167]. The remaining aqueous solution was desalted with Amberlite XAD-7 (6 h at room temperature). The beads were filtered, washed extensively with water and soaked in anhydrous ethanol (100 mL) for 18 h at 4 °C. Finally, the filtrate was reduced to dryness *in vacuo,* yielding a light brown solid (137.5 mg). All extracts were immediately stored at -80 °C.

Aliquots (1 mg) of the extracts were dissolved in 10% acetonitrile:aqueous FA (0.1%; 1 mL) and subjected to low resolution-LC-MS analysis (Gemini C18 column, 0.2 mL/min, gradient from 10-95% acetonitrile:aqueous FA (0.1%) over 35 min). For HR-LC-MS analysis 2 µl of samples (20 µg/mL) were loaded on a C18 trap column (Acclaim, Pepmap 100, 75 µm x 2cm, nanoviper, C18, 3 µm, 100 Å). The trap column was switched online to the analytical column (Thermo, Easy Spray column, PepMap C18 column RSLC, C18, 2 µm, 100 Å, 50 µm x 15cm) after desalting. Samples were eluted over 30min from 5% B to 90% B (Buffer A: 0.1% formic acid, 99.9% milliQ water, Buffer B: 0.1% formic acid, 99.9% acetonitrile) at a flow rate of 300 nL/min and analysed.

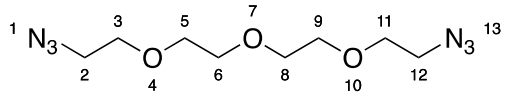
5.3 Linkers and reagents

5.3.1 General remarks

For synthetic procedures, the following solvents were distilled prior to use: diethyl ether under nitrogen from sodium wire, ethyl acetate and petroleum ether from anhydrous potassium carbonate, THF under nitrogen from sodium/potassium with benzophenone as indicator. *tert*-Butyl 35-amino-3,6,9,12,15,18,21,24,27,30,33-undecaoxapentatriacontyl carbamate and 35-azido-3,6,9,12,15,18,21,24,27,30,33-undecaoxapentatriacontan-1-amine were purchased from Polypure, Sweden. Artemisinin and daptomycin were purchased from OChem Incorporation (USA) and used without any further purification. 8-hydroxymanzamine was a kind gift from Prof. Mark T. Hamann (University of Mississippi, USA). TEG was obtained from Sigma (Australia), dried by repeated addition and evaporation of toluene and stored over activated 4Å molecular sieves.
Small-scale distillations were performed by Kugelrohr (Büchi, Switzerland). Water was purified using a Milli-Q Ultrapore Water Purification System (Millipore, USA). Predried solvents (THF, diethylether, DCM) were stored in a PureSolv™ solvent purification system (Innovative Technology, USA), and dispensed under nitrogen as required.

5.3.2 Biotinylated poly(ethyleneglycol) (PEG) linkers and reagents

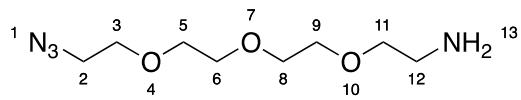
5.3.2.1 TEG-diazide (74; 1-azido-2-(2-(2-(2-azidoethoxy)ethoxy)ethoxy)ethane)



TEA (27.3 mL, 195.6 mmol) was slowly added at 0 °C under nitrogen to a solution of TEG (10.0 g, 51.5 mmol) in anhydrous THF (40 mL). To this, methanesulfonyl chloride (15.1 mL, 195.6 mmol) was added over 30 minutes under nitrogen, resulting in precipitation of a grey solid from a yellow solution. Stirring was continued at 0 °C for one more hour. After stirring for another two hours at room temperature, the solution was chilled on ice/water bath and water (25 mL) added. The reaction was quenched with saturated sodium bicarbonate (15 mL) and the pH adjusted to 8 with sodium hydroxide (2.5 M). Approximately half the volume of THF was evaporated *in vacuo*, sodium azide (6.86 g, 105.6 mmol) added, and the reaction mixture refluxed overnight. The aqueous solution was extracted with diethyl ether (5 × 20 mL), the combined organic layers backwashed with brine (20 mL) and dried over anhydrous magnesium sulfate to yield a yellow oil. This oil was purified by flash chromatography over silica gel, with the column equilibrated in 12.5% ethyl acetate:petroleum ether. A gradient of 12.5-50% ethyl acetate:petroleum ether yielded TEG-diazide (**74**) as a light yellow oil (8.7 g, 69%).

^{1}H-NMR (400 MHz, $CDCl_3$) δ 3.82-3.44 (m, $(H\text{-}3)_2$ – $(H\text{-}11)_2$), 3.37 (t, J = 8.0 Hz, $(H\text{-}2)_2$ and $(H\text{-}12)_2$). Mass spectrum (ESI+) *m/z*: 267 ($[M+Na]^+$), 262 ($[M+H_2O]^+$), 245 ($[M+H]^+$), 239. ν_{max} (Neat film) 2872 (m), 2251 (s), 2108 (s), 1470 (m), 1346 (s), 1299 (m), 1124 (br), 907 (s), 743 (m), 649 (s).

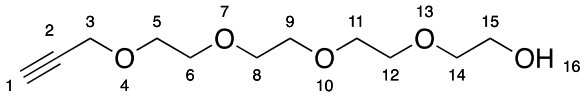
5.3.2.2 N_3-TEG-NH_2 (75; 2-(2-(2-(2-azidoethoxy)ethoxy)ethoxy)ethanamine)



TEG-diazide (**74**; 1.0 g, 4.1 mmol) was dissolved in hydrochloric acid (1 M; 3 mL). With good stirring, triphenyl phosphine (1.18 g, 4.5 mmol) in ethylacetate (15 mL) was added slowly and stirred for a further 14 h. The organic layer was decanted and extracted with aqueous hydrochloric acid (1M; 2 × 3 mL). The combined aqueous layers were saturated with salt, made basic (to pH = 11-12) by addition of sodium hydroxide pellets and thereafter extracted with toluene (3 × 10 mL). The combined

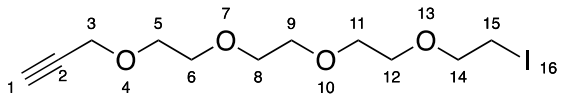
organic layers were washed with brine and dried over potassium hydroxide pellets. The solvents were removed *in vacuo*, leaving a yellow oily residue. This residue was purified by flash chromatography over silica gel, with the column equilibrated in DCM. A gradient of 0-10% methanol:DCM yielded N_3-TEG-NH_2 (**75**) as a light yellow oil (5.5g, 61%). ^{1}H-NMR (400 MHz, $CDCl_3$) δ 3.66 – 3.56 (m, $(H\text{-}3)_2$ – $(H\text{-}11)_2$), 3.49 (t, J = 10.4 Hz, $(H\text{-}2)_2$), 3.35 (t, J = 8.0 Hz, $(H\text{-}12)_2$), 2.85 (b, $(H\text{-}13)_2$). Mass spectrum (ESI+) *m/z:* 241 ($[M+Na]^+$), 219 ($[M+H]^+$).

5.3.2.3 Acetylene-TEG-OH (76; 3,6,9,12-tetraoxapentadec-14-yn-1-ol)



To a solution of predried TEG (10.00 g, 51.5 mmol) in anhydrous THF (50 mL), sodium hydride (1.36 g, 56.6 mmol) was added slowly under nitrogen at 0 °C and stirred for 30 min. Freshly distilled propargyl bromide (4.59 mL, 51.5 mmol) was added at 0 °C over the next 30 minutes, kept on ice/water for another 2 hours and stirred at room temperature overnight. The precipitate was filtered off, the solvent evaporated *in vacuo* and the yellow residue applied to flash chromatography over silica gel. The column was equilibrated to 99:1 DCM:methanol. The residue was dissolved in the initial mobile phase and a gradient of 1-10% methanol:DCM was developed yielding acetylene-TEG-OH (**76**) as a light yellow oil (8.01 g, 67.1%). ^{1}H-NMR (400 MHz, $CDCl_3$) δ 4.11 (d, J = 2.4 Hz, $(H\text{-}3)_2$), 3.64 – 3.54 (m, $(H\text{-}5)_2$ – $(H\text{-}15)_2$), 2.38 (t, J = 3.6 Hz, H-1). Mass spectrum (ESI+) *m/z:* 254 ($[M+Na]^+$), 233 ($[M+H]^+$).

5.3.2.4 Acetylene-TEG-I (77; 1-iodo-3,6,9,12-tetraoxapentadec-14-yne)*



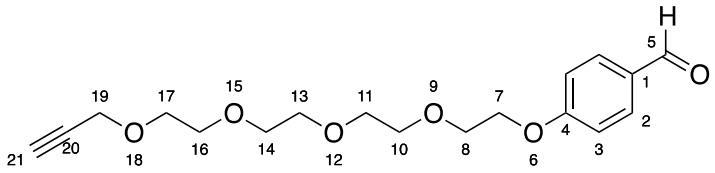
A solution of alcohol (**76**; 200 mg, 0.86 mmol) in dry DCM (1 mL) was added to a solution of iodine (120 mg, 0.95 mmol) and imidazole (64.5 mg, 0.95 mmol) in dry DCM (1 mL) containing a suspension of polymer-bound triphenylphosphine (PPh_3) (320 mg resin equivalent to 0.95 mmol PPh_3) at 0 °C. The reaction mixture was allowed to warm to room temperature, while stirring was continued overnight. The polymer-bound PPh_3 was filtered off, the reaction mixture washed consecutively with

* denotes a new compound

an aqueous solution of saturated sodium thiosulfate, water and brine and then dried over anhydrous magnesium sulfate. The solvent was reduced to dryness *in vacuo*. A silica column was equilibrated in 7:3 petroleum ether:ethyl acetate, developed with a gradient of 30-100% ethyl acetate:petroleum ether yielding acetylene-TEG-I (**77**) as a yellow oil (174 mg, 59%).

^{1}H-NMR (400 MHz, $CDCl_3$) δ 4.19 (d, *J* = 2.4 Hz, $(H\text{-}3)_2$), 3.55-3.71 (m, $(H\text{-}5)_2$ – $(H\text{-}14)_2$), 3.24 (t, *J* = 6.9 Hz, $(H\text{-}15)_2$), 2.41 (t, *J* = 2.4 Hz, H-1). ^{13}C-NMR (100 MHz, $CDCl_3$) δ 80.0 (C-2), 74.9 (C-1), 72.9-69.5 (C-5 – C-14), 62.1 (C-15), 58.8 (C-3).
Mass spectrum (ESI+) *m/z:* 365 ($[M+Na]^+$), 360 ($[M+H_2O]^+$), 343 ($[M+H]^+$). (HRESI+) Found *m/z*: 365.0223, $C_{11}H_{19}IO_4Na$ requires 365.0220.

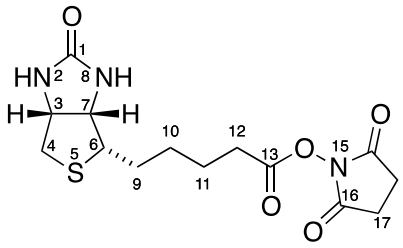
5.3.2.5 Acetylene-TEG-BA (78; 4-(3,6,9,12-tetraoxapentadec-14-ynyloxy)benzaldehyde)*



4-Hydroxy-benzaldehyde (28 mg, 0.23 mmol) was dissolved in anhydrous acetonitrile (2 mL), added to cesium carbonate (112 mg, 0.34 mmol) and stirred for 15 min at room temperature. A solution of "acetylene-TEG-I" (**77**; 117 mg, 0.34 mmol) in anhydrous MeCN (2 mL) was added and the stirring reaction mixture was reduced to 1 mL under nitrogen. After 72 h it was reduced to dryness under nitrogen. The residue was taken up in chloroform and loaded onto a silica column equilibrated in 99:1 chloroform:methanol. The column was developed with a gradient of 1-5% methanol:chloroform and 20 fractions were collected. Fractions 7-12 were combined (TLC) and reduced to dryness *in vacuo* yielding acetylene-TEG-BA (**78**) as a light-yellow oil (74.2 mg, 96.4%).

^{1}H-NMR (400 MHz, $CDCl_3$) δ 9.86 (s, H-5), 7.81 (dt, *J* = 9.5 Hz, 2.4 Hz, $(H\text{-}3)_2$), 7.00 (dt, *J* = 9.5 Hz, 2.3 Hz, $(H\text{-}2)_2$), 4.22-4.15 (m, $(H\text{-}7)_2$ and $(H\text{-}19)_2$), 3.90-3.61 (m, $(H\text{-}8$ – $H\text{-}17)_2$), 2.41 (t, *J* = 2.4 Hz, H-21). ^{13}C-NMR (100 MHz, $CDCl_3$) δ 190.8 (C-5), 163.9 (C-1), 132.0 (C-3), 130.1 (C-2), 114.9 (C-4), 79.7 (C-20), 74.6 (C-21), 70.9-67.8 (8C, C-7 – C-17), 58.4 (C-19). Mass spectrum (ESI+) *m/z:* 359 ($[M+Na]^+$), 337 ($[M+H]^+$). (HRESI+) Found *m/z*: 359.1467, $C_{18}H_{24}O_6Na$ requires 359.1465.

5.3.2.6 Biotin N-hydroxysuccinimide ester (81)

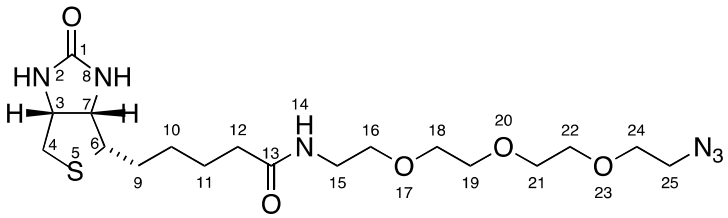


A solution of biotin (200 mg, 0.80 mmol), NHS (92 mg, 0.80 mmol) and DCC (168 mg, 0.80 mmol) in THF:DMSO (95:5; 20 mL) was stirred at room temperature for 18 h. The THF was then removed under a stream of nitrogen and the DMSO by lyophilisation. The residue was recrystallised from isopropanol, yielding biotin-NHS (**81**) as a white powder (235 mg, 88%). M.p. 201-202 °C (Lit.[391] 200-202 °C).

^{1}H-NMR (400 MHz, DMSO-d_6) δ 6.41 (s, H-8), 6.35 (s, H-2), 4.29 (m, H-3), 4.14 (m, H-7), 3.10 (m, H-6), 2.82 (dd, *J* = 12.8 Hz, 5.0 Hz, H-4β), 2.80 (s, $(H\text{-}16)_4$), 2.66 (t, *J* = 7.3 Hz, $(H\text{-}12)_2$), 2.57 (d, *J* = 11.4 Hz, H-4α), 1.40-1.68 (m, $(H\text{-}9 - H\text{-}11)_2$).

Mass spectrum (ESI+) *m/z:* 364 ($[M+Na]^+$), 342 ($[M+H]^+$).

5.3.2.7 Biotin-TEG-N_3 (82; azidoethyl-di(ethylene glycol) ethylamino biotin)

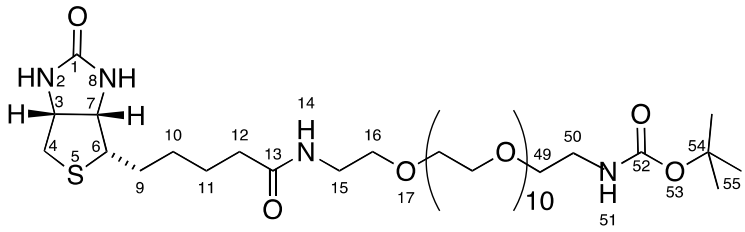


A solution of NH_2-TEG-N_3 (**75**; 70.3 mg, 0.32 mmol) in anhydrous acetonitrile (2.0 mL) was added to a solution of biotin *N*-hydroxysuccinimide ester (**81**; 100 mg, 0.29 mmol) and TEA (82 µL, 0.59 mmol) in DMSO (0.2 mL). The reaction mixture was stirred for 18 h at 25 °C and reduced to dryness under a stream of nitrogen. The residue was taken up in chloroform and loaded onto a silica column equilibrated in 99:1 chloroform:methanol. The column was developed with a gradient of 1-20% methanol:chloroform and 60 fractions were collected. Fractions 44-57 were combined (TLC) and reduced to dryness *in vacuo* yielding Biotin-TEG-N_3 (**82**) as a white sticky powder (102 mg, 78.5%).

^{1}H-NMR (400 MHz, DMSO-d_6) δ 7.81 (t, *J* = 5.4 Hz, H-14), 6.41 (s, H-8), 6.35 (s, H-2), 4.30 (m, H-3), 4.14 (m, H-7), 3.68 – 3.10 (m, $(H\text{-}15)_2$ – $(H\text{-}25)_2$), 3.20 (m, H-6), 2.82 (dd, *J* = 12.4 Hz, 5.1 Hz, H-4β), 2.57 (d, *J* = 12.6 Hz, H-4α), 2.06 (t, *J* = 7.4 Hz, $(H\text{-}12)_2$),

1.68 – 1.40 (m, $(H\text{-}9)_2$ – $(H\text{-}11)_2$). Mass spectrum (ESI+) *m/z*: 467 ($[M+Na]^+$), 445 ($[M+H]^+$).

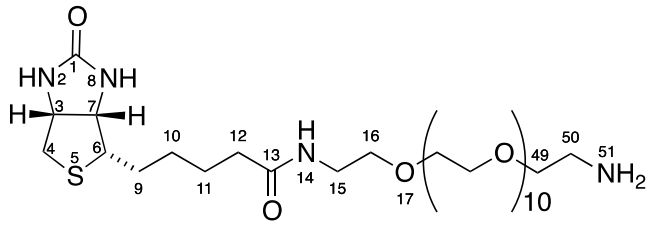
5.3.2.8 Biotin-PEG-NH-Boc (83)



A solution of *tert*-butyl 35-amino-3,6,9,12,15,18,21,24,27,30,33-undecaoxapentatriacontylcarbamate (**79**; *"H_2N-PEG-NH-Boc"*) (40.0 mg, 68.2 μmol) in anhydrous acetonitrile (2.0 mL) was added to a solution of biotin-NHS ester (**81**; 24 mg, 68.2 μmol) and TEA (14.3 μL, 102.3 μmol) in DMSO (0.2 mL). The reaction mixture was stirred for 18 h at 25 °C and reduced to dryness under a stream of nitrogen. The residue was taken up in chloroform and loaded onto a short silica column (5 cm) equilibrated in 9:1 ethyl acetate:methanol. The column was developed with a gradient of 10-50% ethyl acetate:methanol yielding amide **83** as a colourless oil (39 mg, 72%).

^{1}H-NMR (400 MHz, $CDCl_3$) δ 6.84 (t, *J* = 5.4 Hz, H-14), 6.60 (bs, H-8), 5.73 (bs, H-2), 5.06 (bs, H-51), 4.46 (m, H-3), 4.27 (m, H-7), 3.80 – 3.04 (m, $(H\text{-}16)_2$ – $(H\text{-}50)_2$), 3.37 (m, $(H\text{-}15)_2$), 3.10 (m, H-6), 2.88 (dd, *J* = 12.7 Hz, 4.8 Hz, H-4β), 2.70 (d, *J* = 12.8 Hz, H-4α), 2.18 (t, *J* = 7.5 Hz, $(H\text{-}12)_2$), 1.79 – 1.33 (m, $(H\text{-}9)_2$ – $(H\text{-}11)_2$), 1.40 (s, t-butyl). Mass spectrum (ESI+) *m/z*: 872 ($[M+H]^+$), 772 ($[M\text{-}Boc+H]^+$), 386 ($[M\text{-}Boc+2H]^{2+}$).

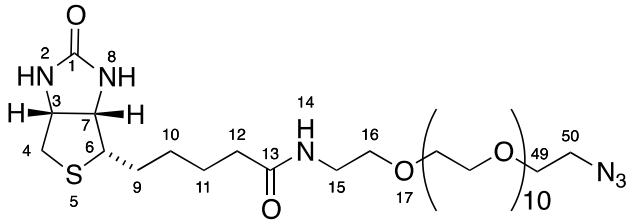
5.3.2.9 Biotin-PEG-NH_2 (84)



Biotin-PEG-NH-Boc (**83**; 39 mg, 44.3 μmol) was stirred in DCM:TFA (55%; 1 mL) for 30 min. TFA and DCM were removed under a stream of nitrogen. The residue was redissolved in methanol and stirred over Amberlyst A21 cation exchange resin (hydroxide form) for 30 min; the resin was filtered off and the solvent was reduced to dryness *in vacuo* to give the free amine (**84**) as a colourless oil (34.1 mg, 100%).

^{1}H-NMR (400 MHz, DMSO-d_6) δ 7.90 (bs, (H-51$_2$), 7.85 (t, *J* = 5.4 Hz, H-14), 6.40 (s, H-8), 6.36 (s, H-2), 4.30 (m, H-3), 4.12 (m, H-7), 3.70 – 3.04 (m, (H-16)$_2$ – (H-50)$_2$), 3.09 (m, H-6), 2.96 (m, (H-15)$_2$), 2.81 (dd, *J* = 12.7 Hz, 4.8 Hz, H-4β), 2.58 (d, *J* = 12.8 Hz, H-4α), 2.05 (t, *J* = 7.5 Hz, (H-12)$_2$), 1.69 – 1.18 (m, (H-9)$_2$ – (H-11)$_2$). Mass spectrum (ESI+) *m/z:* 772 ([M+H]$^+$), 386 ([M+2H]$^{2+}$).

5.3.2.10 Biotin-PEG-N$_3$* (86)

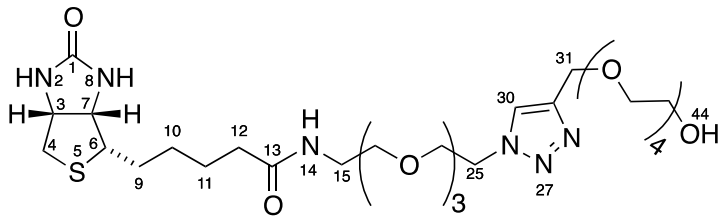


To a cooled (0 °C) mixture of biotin (119.9 mg, 0.49 mmol), EDC (141.1 mg, 0.73 mmol), and HOBt (112.6 mg, 0.73 mmol) in anhydrous DMF (5 mL) was added O-(2-aminoethyl)-O′-(2-azidoethyl)decaethylene glycol (**85**; 140.0 mg, 0.245 mmol) in DMF (1 mL) dropwise. The reaction mixture was stirred at 0 °C for 0.5 h and was then allowed to warm to room temperature. The mixture was then stirred for a further 24 h, diluted with water (100 mL) and passed through a disposable C18 cartridge (2000 mg bed weight). The column was washed with water (50 mL) to remove the DMF and the product eluted with methanol. The mixture was reduced to dryness by rotary evaporation and the residue taken up in chloroform and loaded onto a silica column equilibrated in 95:5 chloroform:methanol. The column was developed with a gradient of 5-15% methanol:chloroform yielding the azide **86** as a colourless oil (151.4 mg, 77.4%).

^{1}H-NMR (400 MHz, $CDCl_3$) δ 6.85 (t, *J* = 5.3 Hz, H-14), 6.54 (bs, H-8), 5.68 (bs, H-2), 4.47 (m, H-3), 4.28 (m, H-7), 3.70 – 3.08 (m, (H-15)$_2$ – (H-50)$_2$), 2.87 (dd, *J* = 12.7 Hz, 4.8 Hz, H-4β), 2.71 (d, H-4α), 2.20 (t, *J* = 7.4 Hz, (H-12)$_2$), 1.76 – 1.35 (m, (H-9)$_2$ – (H-11)$_2$). ^{13}C-NMR (100 MHz, $CDCl_3$) δ 173.2 (C-13), 163.9 (C-1), 70.9 – 69.3 (C-16 – C-49), 61.8 (C-7), 60.1 (C-3), 55.5 (C-6), 50.2 (C-50), 40.6 (C-4), 39.2 (C-15), 35.9 (C-12), 28.2 (C-10), 28.1 (C-9), 25.5 (C-11). Mass spectrum (ESI+) *m/z:* 820 ([M+Na]$^+$), 798 ([M+H]$^+$). (HRESI+) Found *m/z*: 819.4163, $C_{34}H_{64}N_6O_{13}SNa$ requires 819.4144.

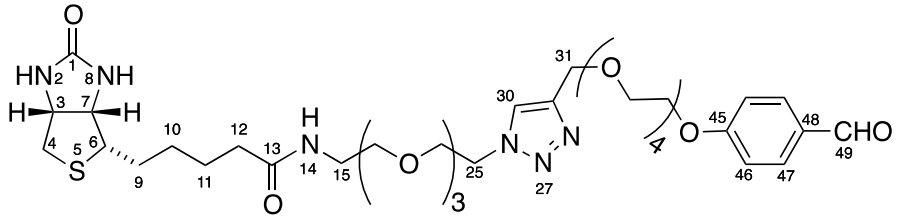
* denotes a new compound

5.3.2.11 Biotin-TEG-triazole-TEG-OH[*] (87)



Copper(II) sulfate (0.1 mM; 0.25 mL, 23 µmol) was added to a solution of sodium ascorbate (9.1 mg, 46 µmol) in *tert*-butanol (0.25 mL) and stirred until the initially forming brown precipitate turned to a yellow solution (5 min). To this, a solution of Biotin-TEG-N_3 (**82**; 250 mg, 0.56 mmol) and compound acetylene-TEG-OH (**76;** 130 mg, 0.56 mmol) in *tert*-butanol:water (1:1; 1 mL) was added and stirred for 24 h. The reaction mixture was reduced under nitrogen and freeze-dried. The residue was suspended in 98:2 chloroform:methanol and loaded onto a silica column equilibrated in 99:1 chloroform:methanol. A gradient of 1-20% methanol:chloroform was applied yielding Biotin-TEG-triazole-TEG-OH (**87**) as a light-yellow oil (290.3 mg, 76.1%). ^{1}H-NMR (400 MHz, DMSO-d_6) δ 8.03 (s, H-30), 7.80 (t, *J* = 5.4 Hz, H-14), 6.40 (s, H-8), 6.34 (s, H-2), 4.51 (s, $(H\text{-}31)_2$), 4.49 (t, *J* = 5.1 Hz, $(H\text{-}25)_2$), 4.29 (dd, *J* = 7.7 Hz, 5.0 Hz, H-3), 4.12 (m, H-7), 3.80 (t, *J* = 5.3 Hz, $(H\text{-}24)_2$), 3.63 – 3.26 (m, $(H\text{-}16)_2$ – $(H\text{-}23)_2$ and $(H\text{-}33)_2$ – $(H\text{-}42)_2$), 3.17 (m, $(H\text{-}15)_2$), 3.08 (m, H-6), 2.81 (dd, *J* = 12.7 Hz, 4.8 Hz, H-4β), 2.56 (d, *J* = 12.6 Hz, H-4α), 2.05 (t, *J* = 7.5 Hz, $(H\text{-}12)_2$), 1.60 (m, H-9a), 1.49 (m, $(H\text{-}11)_2$), 1.45 (m, H-9b), 1.31 (m, $(H\text{-}10)_2$). ^{13}C-NMR (100 MHz, $CDCl_3$) δ 174.6 (C-13), 162.7 (C-1), 143.8 (C-29), 124.2 (C-30), 72.3 – 69.5 (C-16 – C-24 and C-33 – C-42), 68.8 (C-24), 63.5 (C-31), 61.0 (C-7), 60.2 (C-43), 59.2 (C-3), 55.4 (C-6), 49.3 (C-25), 39.8 (C-4), 38.4 (C-15), 35.1 (C-12), 28.2 (C-10), 28.1 (C-9), 25.3 (C-11). Mass spectrum (ESI+) *m/z:* 700 ($[M+Na]^+$), 678 ($[M+H]^+$), 339 ($[M+2H]^{2+}$). (HRESI+) Found *m/z*: 699.3370, $C_{29}H_{52}N_6O_{10}S$ requires 699.3358.

5.3.2.12 Biotin-TEG-triazole-TEG-BA (88)

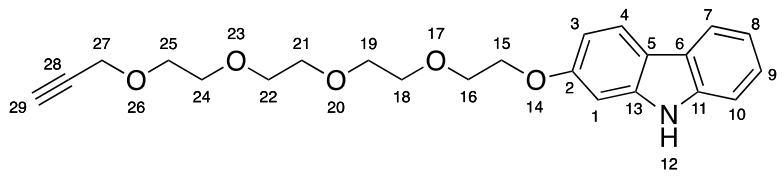


[*] denotes a new compound

Copper(II) sulfate (0.1 mM; 0.25 mL, 23 µmol) was added to a solution of sodium ascorbate (9.1 mg, 46 µmol) in *tert*-butanol (0.25 mL) and stirred until the initially forming brown precipitate turned to a yellow solution (5 min). To this, a solution of azide (**82**; 34 mg, 76.5 µmol) and acetylene (**78**; 30.8 mg, 91.8 µmol) in *tert*-butanol:water (1:1, 1 mL) was added and stirred for 18 h. The reaction mixture was reduced under nitrogen and freeze-dried. The residue was suspended in 98:2 chloroform:methanol and loaded onto a silica column equilibrated in 99:1 chloroform:methanol. A gradient of 1-20% chloroform:methanol was applied yielding the triazole **88** as a light-brown oil (48.4 mg, 81.0%).

^{1}H-NMR (400 MHz, $CDCl_3$) δ 9.83 (s, H-49), 7.78 (dt, *J* = 9.5 Hz, 2.3 Hz, $(H\text{-}46)_2$), 7.72 (s, H-30), 6.98 (dt, *J* = 9.5 Hz, 2.2 Hz, $(H\text{-}47)_2$), 6.87 (t, *J* = 5.4 Hz, H-14), 6.43 (bs, H-8), 5.70 (bs, H-2), 4.62 (s, $(H\text{-}31)_2$), 4.49 (t, *J* = 5.1 Hz, $(H\text{-}25)_2$), 4.43 (m, H-3), 4.25 (m, H-7), 4.17 (t, *J* = 4.7 Hz, $(H\text{-}43)_2$), 3.83 – 3.46 (m, $(H\text{-}16)_2$ – $(H\text{-}24)_2$ and $(H\text{-}33)_2$ – $(H\text{-}42)_2$), 3.37 (m, $(H\text{-}15)_2$), 3.08 (m, H-6), 2.84 (dd, *J* = 12.7 Hz, 4.8 Hz, H-4β), 2.68 (d, *J* = 12.8 Hz, H-4α), 2.15 (t, *J* = 7.5 Hz, $(H\text{-}12)_2$), 1.73 – 1.33 (m, $(H\text{-}9)_2$ – $(H\text{-}11)_2$). ^{13}C-NMR (100 MHz, $CDCl_3$) δ 190.9 (C-49), 173.4 (C-13), 164.1 (C-48), 163.9 (C-1), 144.8 (C-29), 132.0 (C-46), 130.0 (C-45), 123.9 (C-30), 114.9 (C-47), 70.9 – 69.5 (C-16 – C-24 and C-33 – C42), 67.8 (C-43), 64.5 (C-31), 61.8 (C-7), 60.2 (C-3), 55.7 (C-6), 50.2 (C-25), 40.5 (C-4), 39.1 (C-15), 35.9 (C-12), 28.3 (C-10), 28.1 (C-9), 25.6 (C-11). Mass spectrum (ESI+) *m/z:* 803 ($[M+Na]^+$). (HRESI+) Found *m/z*: 803.3640, $C_{36}H_{56}N_6O_{11}SNa$ requires 803.3620.

5.3.2.13 Acetylene-TEG-carbazole (96; 2-(3,6,9,12-tetraoxapentadec-14-yn-1-yloxy)-9H-carbazole)*

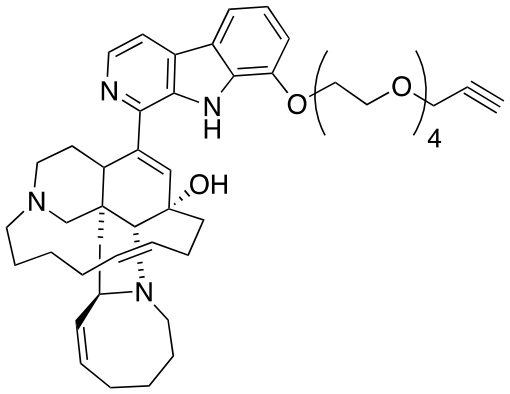


A solution of acetylene-TEG-I (**77**; 10.0 mg, 29.2 µmol) in anhydrous DMF (500 µL) was added to 2-hydroxycarbazole (5.9 mg, 32.2 µmol) and cesium carbonate (14.3 mg, 43.8 µmol) under nitrogen. The reaction mixture was stirred at room temperature for 60 min, then heated to 60 °C for 5 min and left stirring at room temperature overnight. Distilled water (4.5 mL) was added and the entire mixture

* denotes a new compound

was put through a preconditioned Strata-X cartridge (100 mg bed weight). After extensive washing with distilled water, the organic compounds were eluted with methanol (5 mL) and the solvent evaporated *in vacuo* to yield a dark brown oily residue. This residue was purified by flash chromatography over silica gel, with the column equilibrated in 99:1 chloroform:methanol. A gradient of 1-10% methanol:chloroform was applied to yield **96** as a light brown oil (11.1 mg, 87%). ^{1}H-NMR (400 MHz, $CDCl_3$) δ 8.23 (s, H-12), 7.96 (d, *J* = 7.8 Hz, H-7), 7.91 (d, *J* = 8.6 Hz, H-3), 7.38 (d, *J* = 8.0 Hz, H-10), 7.32 (dd, *J* = 7.5 Hz, 1.1 Hz, H-9), 7.19 (dd, *J* = 7.4 Hz, 1.0 Hz, H- 8), 6.99 (d, *J* = 2.1 Hz, H-1), 6.85 (dd, *J* = 8.5 Hz, 2.2 Hz, H-4), 4.23 (dd, *J* = 5.4 Hz, 4.3 Hz, $(H\text{-}15)_2$), 4.18 (d, *J* = 2.2 Hz, $(H\text{-}27)_2$), 3.67-3.70 (m, $(H\text{-}16)_2$ – $(H\text{-}25)_2$), 2.41 (t, *J* = 2.4 Hz, H-29). ^{13}C-NMR (100 MHz, $CDCl_3$) δ 158.2 (C-2), 140.8 (C-13), 139.6 (C-11), 124.6 (C-9), 123.5 (q, C-6), 121.0 (C-3), 119.5 (C-7), 119.4 (C-8), 117.4 (q, C-5), 110.4 (C-10), 108.8 (C-4), 96.0 (C-1), 79.6 (C-28), 74.7 (C-29), 70.9-69.1 (C-16 – C-25), 68.0 (C-15), 58.4 (C-27). Mass spectrum (ESI+) *m/z:* 420 ($[M+H_2O]^+$), 398 ($[M+H]^+$). (HRESI+) Found *m/z*: 398.1959, $C_{23}H_{28}NO_5$ requires 398.1962.

5.3.2.14 Manzamine-TEG-acetylene (97)*

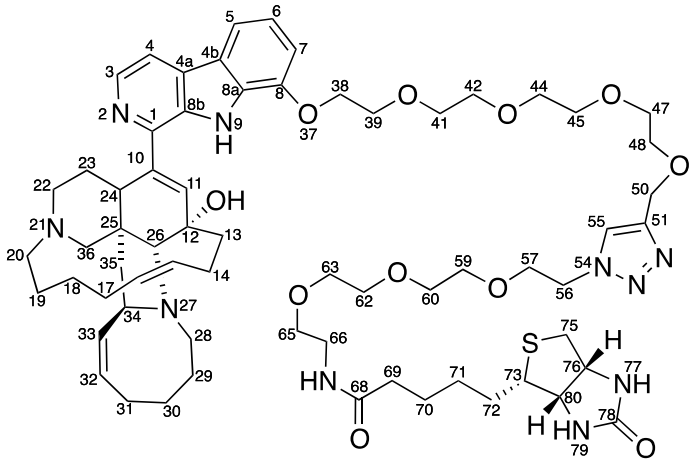


Manzamine hydrochloride (4.20 mg, 7.0 μmol) was dissolved in dry DMF (400 μL), added to cesium carbonate (3.42 mg, 10.5 μmol) at 0 °C and stirred for 30 min at 0 °C. Acetylene-TEG-I (**77**; 2.40 mg, 7.0 μmol) was added in DMF (100 μL) at 0 °C and the reaction mixture was kept stirring while allowed to warm to ambient temperature (30 min). It was then heated at 50 °C. An aliquot (2 μL) was taken under inert conditions every 20 min to monitor the reaction. The aliquot was diluted in dist. water (25 μL) and analysed by HPLC (analytical C18, 1mL/min, 5-65% acetonitrile:aqueous TFA (0.1%) over 15 min). After 2 h the mixture was cooled on

* denotes a new compound

ice (10 min) and concentrated under a stream of nitrogen, diluted with water (3 mL) and loaded onto a disposable C18 cartridge (200 mg bed weight). The column was washed with 5:95 methanol:water (3 mL) to remove the DMF and the product eluted with 1:1 methanol:aqueous TFA (0.1%; 2 mL). TFA and methanol were removed under a stream of nitrogen and the resulting aqueous solution subjected to HPLC separation (semi-preparative C18 column, 4.73 mL/min, gradient from 5-65% acetonitrile:aqueous TFA (0.1%) over 15 min). The peak eluting after 11.4 min was collected and freeze-dried to yield the TFA salt of manzamine-TEG-acetylene (**97**) as colourless solid (4.18 mg, 66.8%). Mass spectrum (ESI+) *m/z:* 802 ($[M+Na]^+$), 390 ($[M+2H]^{2+}$). (HRESI+) Found *m/z*: 801.4575, $C_{47}H_{62}N_6O_{11}SNa$ requires 801.4562. The solid was used immediately for the next step.

5.3.2.15 Biotin-manzamine (98)*



The TFA salt of the Manzamine-TEG-acetylene (**97**; 4.2 mg, 4.62 µmol) was dissolved in absolute methanol (2 mL) and stirred over Amberlyst A21 cation exchange resin (hydroxide form) for 30 min. The resin was filtered off and washed with methanol (2 mL). The organic layers were combined and the solvent removed *in vacuo* to yield the free base of the acetylene (**97**; 3.6 mg, 4.62 µmol, 100%), which was immediately dissolved in 1:1 ethanol:water (300 µL) and added to a mixture of biotin-TEG-azide (**82**; 100 µL of 22.6 mg/mL ethanol, 5.08 µmol), sodium ascorbate (100 µL of 3.7 mg/mL water, 1.85 µmol), copper(II) sulfate (9.2 µL of 100mM in water, 0.92 µmol) and Tris-[(1-benzyl-1H-1,2,3-triazol-4-yl) methyl]amine (TBTA; 10 µL of 2.3 mM in ethanol, 0.23 µmol) and kept stirring at room temperature for 3 h. The reaction mixture was directly loaded onto a disposable C18 cartridge (200 mg bed weight) and

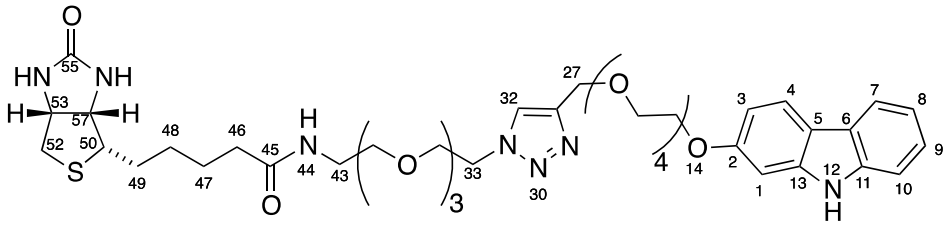
* denotes a new compound

washed with 5:95 methanol:water (2 mL). The product was eluted with 80:20 acetonitrile:aqueous TFA (0.1%; 2 mL) and concentrated under a stream of nitrogen. The aqueous reaction mixture was separated by HPLC (preparative C18, 9.99 mL/min, with a gradient of 15-60% acetonitrile:aqueous TFA (0.1%) over 15 min). The peak eluting after 10.3 min was collected and instantly frozen. The combined fractions were freeze-dried to yield biotin-manzamine (**98**) as colourless solid (4.4 mg, 78%).

^{1}H-NMR (600 MHz, $CDCl_3$) δ 11.07 (s, H-9), 8.34 (d, *J* = 5.1 Hz, H-3), 7.83 (d, *J* = 5.1 Hz, H-4), 7.79 (bs, H-67), 7.78 (s, H-55), 7.66 (d, *J* = 7.9 Hz, H-5), 7.15 (t, *J* = 7.8 Hz, H-6), 7.01 (d, *J* = 7.7 Hz, H-7), 6.70 (m, H-11), 6.31 (m, H-32), 5.58 (m, H-15), 5.55 (m, H-16), 5.45 (t, *J* = 9.8 Hz, H-33), 4.92 (m, H-34), 4.66 (s, H-50), 4.52 (t, *J* = 5.1 Hz, $(H\text{-}56)_2$), 4.47 (m, H-76), 4.42 (m, H-38a), 4.35 (m, H-38b), 4.29 (m, H-80), 4.21 (m, H-39a), 4.07 (m, H-39b), 4.06 (m, H-28a), 3.93 - 3.51 (m, $(H\text{-}41)_2$ - $(H\text{-}48)_2$ and $(H\text{-}59)_2$ - $(H\text{-}63)_2$), 3.86 (t, *J* = 5.1 Hz, $(H\text{-}57)_2$), 3.70 (m, H-26), 3.50 (m, $(H\text{-}65)_2$), 3.42 (m, $(H\text{-}66)_2$), 3.24 (m, H-28b), 3.11 (m, H-73), 2.96 (m, H-36a), 2.96 (m, H-23a), 2.94 (m, H-22a), 2.87 (m, H-75a), 2.71 (m, H-75b), 2.69 (m, H-24), 2.63 (m, H-20a), 2.55 (m, H-29a), 2.49 (m, H-17a), 2.48 (m, H-35a), 2.43 (m, H-36b), 2.43 (m, H-20b), 2.38 (m, H-31a), 2.29 (m, H-31b), 2.23 (m, $(H\text{-}14)_2$), 2.17 (m, $(H\text{-}69)_2$), 2.10 (m, H-13a), 1.73 (m, H-13b), 2.03 (m, H-29b), 2.01 (m, H-30a), 1.93 (m, H-22b), 1.92 (m, H-35b), 1.85 (m, H-19a), 1.74 (m, H-23b), 1.70 (m, H-72a), 1.64 (m, $(H\text{-}70)_2$), 1.62 (m, H-17b), 1.60 (m, H-72b), 1.51 (m, H-18a), 1.50 (m, H-19b), 1.47 (m, H-30b), 1.41 (m, $(H\text{-}71)_2$), 1.21 (m, H-18b). ^{13}C-NMR (150 MHz, $CDCl_3$) δ 172.9 (C-68), 163.2 (C-78), 146.2 (C-8), 144.8 (C-51), 143.1 (C-1), 142.3 (C-32), 141.7 (C-10), 137.6 (C-3), 135.5 (C-11), 133.3 (C-9a), 133.0 (C-16), 132.5 (C-8a), 130.1 (C-4a), 127.0 (C-15), 124.0 (C-33), 124.0 (C-55), 122.8 (C-4b), 120.3 (C-6), 114.1 (C-4), 113.4 (C-5), 109.6 (C-7), 77.7 (C-26), 71.1 (C-12), 70.9 - 69.3 (C-40 - C-48 and C-59 -C-63), 70.5 (C-36), 69.6 (C-39), 69.6 (C-65), 69.2 (C-57), 68.6 (C-38), 64.6 (C-50), 61.8 (C-80), 60.1 (C-76), 57.0 (C-34), 55.1 (C-73), 53.5 (C-20), 53.2 (C-28), 50.1 (C-56), 49.2 (C-22), 47.0 (C-25), 44.6 (C-35), 40.5 (C-24), 40.5 (C-75), 39.2 (C-66), 39.1 (C-13), 35.7 (C-69), 33.8 (C-23), 28.3 (C-31), 27.9 (C-71), 27.8 (C-72), 26.5 (C-29), 26.3 (C-18), 25.4 (C-70), 25.0 (C-17), 24.5 (C-19), 24.5 (C-30), 20.7 (C-14). Mass spectrum (ESI+) *m/z:* 409 ($[M+2Na+H]^{3+}$), 613 ($[M+2Na]^{2+}$). (HRESI+) Found *m/z*: 1223.6886, $C_{65}H_{91}N_{10}O_{11}S$ requires 1223.6819.

The purity of the compound was verified by reinjecting an aliquot (30 µL; 0.41 mM; in 15% acetonitrile) onto an analytical C18 RP-HPLC column (1 mL/min, 15-80% acetonitrile in aqueous TFA (0.05%) over 30min). The trace contained exclusively one peak, which eluted after 16.9 min and showed absorbance maxima at 223 nm, 277 nm and 362 nm.

5.3.2.16 Biotin-TEG-triazole-TEG-carbazole[*] (99)



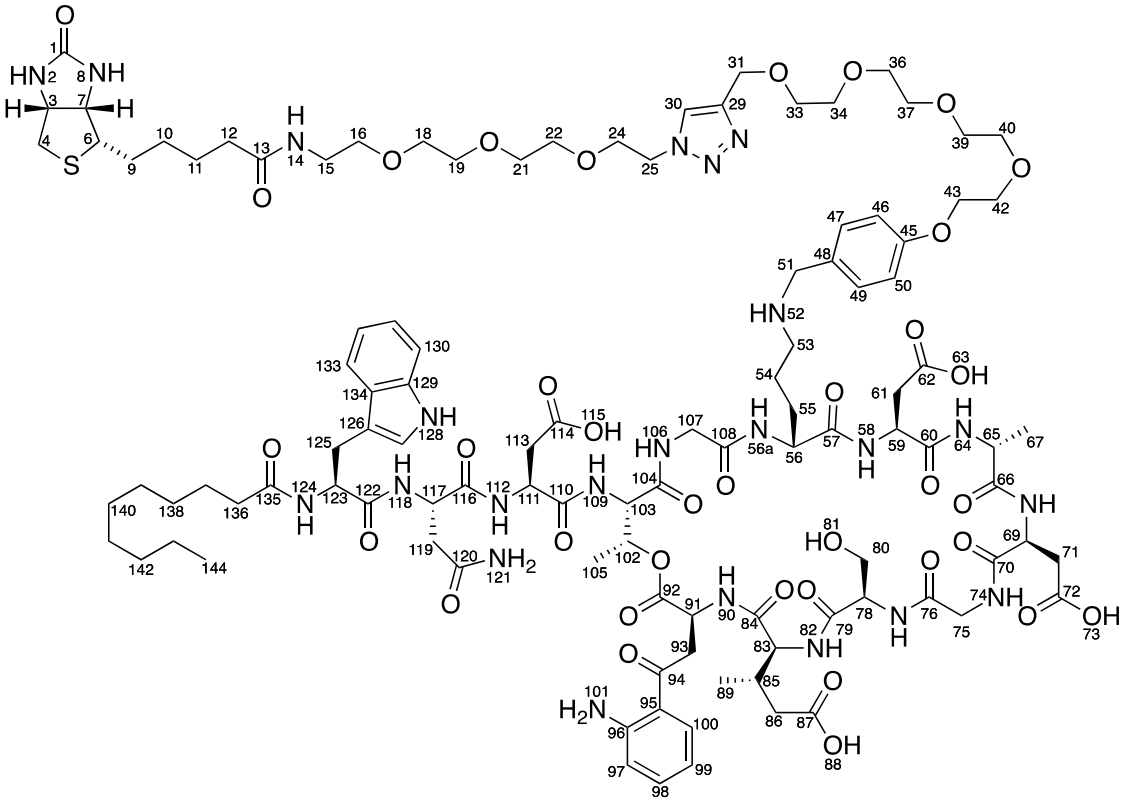
Copper(II) sulfate (0.1 mM; 55.9 µL, 5.6 µmol) was added to a solution of sodium ascorbate (2.2 mg, 11.2 µmol) in ethanol:water (1:1, 1 mL) and stirred until the initially formed brown precipitate turned yellow (5 min). To this, a solution of azide (**82**; 13.7 mg, 30.7 µmol) and acetylene (**96**; 11.1 mg, 27.9 µmol) in ethanol:water (1:1; 1 mL) was added and stirred for 3 h. The reaction mixture was diluted with water (2 mL) and loaded onto a disposable Strata-X cartridge (100 mg bed weight). After washing with 95:5 water:methanol (5 mL) the mixture was eluted with absolute methanol (3 mL). The solvent was removed under a stream of nitrogen and the residue was purified by HPLC (semi-preparative C18 column, 4.7 mL/min, gradient from 30-80% acetonitrile:water over 25 min). The peak eluting after 13.4 min was collected, acetonitrile was removed under a stream of nitrogen and the water lyophilised to give Biotin-TEG-triazole-TEG-carbazole (**99**) as white solid (16.6 mg, 70.6%).

^{1}H-NMR (400 MHz, DMSO-d_6) δ 11.14 (s, H-12), 8.04 (s, H-32), 7.97 (d, J = 7.7 Hz, H-7), 7.95 (d, J = 8.6 Hz, H-3), 7.83 (t, J = 5.4 Hz, H-44), 7.41 (d, J = 8.0 Hz, H-10), 7.27 (dd, J = 7.5 Hz, 1.1 Hz, H-9), 7.19 (dd, J = 7.0 Hz, 1.0 Hz, H- 8), 6.97 (d, J = 2.4 Hz, H-1), 6.76 (dd, J = 8.6 Hz, 2.2 Hz, H-4), 6.40 (s, H-56), 6.35 (s, H-54), 4.50 (s, $(H\text{-}27)_2$), 4.49 (t, J = 5.5 Hz, H-33), 4.29 (m, H-53), 4.16 (dd, J = 5.4 Hz, 4.3 Hz, $(H\text{-}15)_2$), 4.11 (m, H-57), 3.84 - 3.11 (m, $(H\text{-}16)_2$ – $(H\text{-}25)_2$ and $(H\text{-}32)_2$ – $(H\text{-}42)_2$), 3.17 (m, $(H\text{-}43)_2$), 3.07 (m, H-50), 2.80 (dd, J = 12.4 Hz, 5.1 Hz, H-4β), 2.57 (d, J = 12.6 Hz, H-4α), 2.05 (t, J = 7.3 Hz, $(H\text{-}46)_2$), 1.70 – 1.16 (m, $(H\text{-}47)_2$ – $(H\text{-}49)_2$). ^{13}C-NMR (100 MHz, $CDCl_3$) δ

[*] denotes a new compound

172.2 (C-45), 163.2 (C-55), 158.1 (C-2), 144.3 (C-28), 141.6 (C-13), 140.2 (C-11), 124.7 (C-9), 124.6 (C-32), 123.1 (C-6), 121.4 (C-3), 119.7 (C-7), 119.0 (C-8), 116.7 (C-5), 111.1 (C-10), 108.5 (C-4), 95.7 (C-1), 70.4 - 69.2 (C-16 – C-25 and C-34 – C-42), 67.8 (C-15), 64.0 (C-27), 61.5 (C-57), 59.7 (C-53), 55.9 (C-50), 49.8 (C-33), 40.3 (C-52), 38.9 (C-43), 35.6 (C-46), 28.7 (C-48), 28.5 (C-49), 25.7 (C-47). Mass spectrum (ESI+) *m/z*: 842 ($[M+H]^+$), 422 ($[M+Na]^+$). (HRESI+) Found *m/z*: 864.3943, $C_{41}H_{59}N_7O_{10}SNa$ requires 864.3936.

5.3.2.17 Biotin-daptomycin (101)*



A solution of Biotin-TEG-triazole-TEG-benzaldehyde (**88**; 10.0 mg, 12.8 µmol) in dry DMF (200 µL) was added to daptomycin (20.7 mg, 12.8 µmol) under nitrogen and stirred for 1 h. To this, $NaBH(OAc)_3$ (16.3 mg, 76.8 µmol) was added and kept stirring at ambient temperature for 24h. The solvent was removed under a stream of nitrogen. The residue was dissolved in 5:95 methanol:water (1 mL) and separated by HPLC (analytical Vydac C4 250.0 × 4.60, 5µ, 300Å, 1 mL/min, gradient from 7-30% acetonitrile:5mM ammonium phosphate buffer over 30 min). The peak eluting after 20.1 min was collected and freeze-dried. The residue was taken up in 5:95 methanol:water (2 mL) and loaded onto a disposable C18 cartridge (200 mg bed

* denotes a new compound

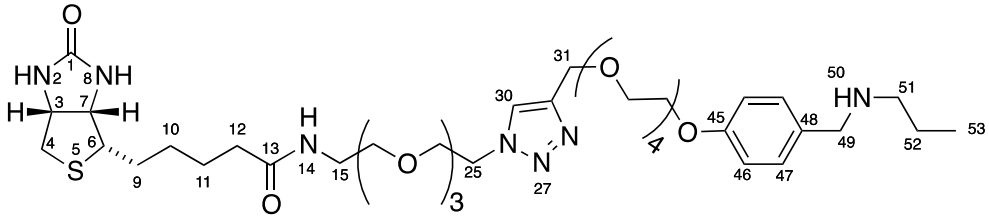
weight). The cartridge was washed with water (3 mL) to remove any salts and the compound eluted with methanol. The solvent was evaporated *in vacuo* yielding biotin-daptomycin (**101**) as light yellow solid (8.4 mg, 27%).

^{1}H-NMR, ^{13}C-NMR, and 2D-NMR experiments are displayed in Appendix 8.2.2.

Mass spectrum (ESI+) m/z: 1193.6 ($[M+2H]^{2+}$), 796.1 ($[M+3H]^{3+}$). (HRESI+) Found m/z: 1215.0312. $C_{108}H_{157}N_{23}O_{36}SNa_2$ requires 1215.0333.

The purity of the biotinylated daptomycin was verified by reinjecting an aliquot (30 µL) of a solution of product **101** (0.5 mg; 0.21 µmol; 1 mL in 20% acetonitrile:5mM ammonium phosphate buffer over 30 min) onto an analytical C18 RP-HPLC column (1 mL flow, 20-70% acetonitrile:5mM ammonium phosphate buffer over 30min). The trace contained exclusively one peak, which eluted after 16.9 min and showed absorbance maxima at 224 nm, 259 nm and 369 nm.

5.3.2.18 Biotin-TEG-triazole-TEG-BA-propylamine (103)*



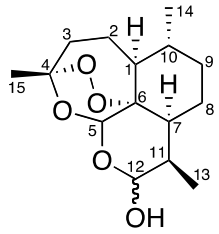
To a solution of aldehyde (**88**; 15.1 mg, 19.3 µmol) in methanol-d_4 (0.60 mL), was added $NaBH(OAc)_3$ (41.0 mg, 193.4 µmol) and 1-propylamine (9.5 µL, 116.0 µmol) and the reaction mixture stirred at ambient temperature for 5 min. ^{1}H-NMR showed the presence of imine [δ_H 8.16 (s, H-49)] and therefore more $NaBH(OAc)_3$ (25.0 mg, 118.0 µmol) was added. After another 0.5 h, NMR revealed the reaction was completed (no imine present) and the solvent was removed under a stream of nitrogen. The residue was suspended in chloroform and loaded onto a short silica column (5 cm) equilibrated in 95:5 chloroform:methanol. The column was developed with a gradient of 5-15% methanol:chloroform yielding Biotin-TEG-triazoleTEG-BA-propylamine (**103**) as a colourless oil (7.2 mg, 45%).

^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.76 (s, H-30), 7.22 (d, *J* = 8.1 Hz, $(H\text{-}47)_2$), 6.85 (d, *J* = 8.5 Hz, $(H\text{-}46)_2$), 6.57 (t, *J* = 5.7 Hz, H-14), 6.02 (bs, H-8), 5.22 (bs, H-2), 4.67 (s, $(H\text{-}31)_2$), 4.53 (t, *J* = 3.4 Hz, $(H\text{-}25)_2$), 4.47 (m, H-3), 4.30 (m, H-7), 4.10 (t, *J* = 3.2 Hz, H-15), 3.87 (t, *J* = 3.35, H-43), 3.83 (t, *J* = 3.2 Hz, H-49), 3.79 – 3.36 (m, $(H\text{-}16)_2$ – $(H\text{-}24)_2$ and (H-

* denotes a new compound

33)$_2$ – (H-42)$_2$), 3.12 (m, H-6), 2.89 (dd, *J* = 12.8 Hz, 4.8 Hz, H-4β), 2.71 (d, *J* = 12.6 Hz, H-4α), 2.59 (t, *J* = 7.3 Hz, (H-51)$_2$), 2.18 (t, *J* = 7.0 Hz, (H-12)$_2$), 1.78 – 1.34 (m, (H-9)$_2$ – (H-11)$_2$), 1.53 (q, *J* = 7.3 Hz, H-52), 0.91 (t, *J* =7.4 Hz, (H-53)$_3$). ^{13}C-NMR (100 MHz, $CDCl_3$) δ 173.2 (C-13), 163.6 (C-1), 157.9 (C-45), 145.0 (C-29), 132.2 (C-48), 129.5 (C-47), 123.9 (C-30), 114.6 (C-46), 70.9 – 69.5 (C-16 – C-24 and C-33 – C-42), 67.5 (C-43), 64.6 (C-31), 61.8 (C-7), 60.1 (C-3), 55.5 (C-6), 53.2 (C-51), 51.0 (C-49), 50.2 (C-25), 40.6 (C-4), 39.2 (C-15), 35.9 (C-12), 28.2 (C-10), 28.1 (C-9) 25.5 (C-11), 22.9 (C-52), 11.8 (C-53). Mass spectrum (ESI+) *m/z:* 824 ([M+H]$^+$), 412 ([M+2H]$^+$). (HRESI+) Found *m/z*: 824.4588, $C_{39}H_{66}N_7O_{10}S$ requires 824.4586.

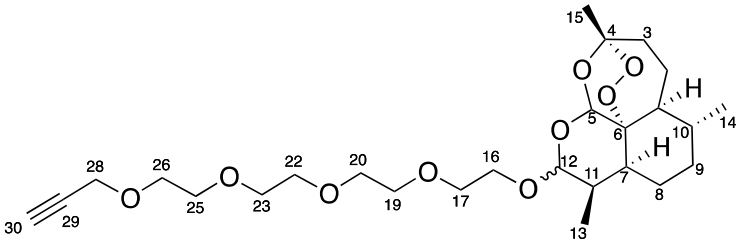
5.3.2.19 Dihydroartemisinin (105; DHA)



Artemisinin (500 mg, 1.8 mmol) was dissolved in absolute methanol (40 mL) and cooled to 0 °C. Granular $NaBH_4$ (250 mg, 6.6 mmol) was added in small portions over 30 min and the solutions was kept stirring at 0-5 °C for 2 h. The reaction mixture was neutralised with 30% acetic acid:methanol (5 mL) and the solvent removed *in vacuo*. The solid residue was extracted with ethyl acetate (2 × 50 mL) and filtered. The organic solvent was again removed *in vacuo* and the residue recrystallised from ethyl acetate:hexane. The crystals were filtered off, washed with hexane and dried *in vacuo* to yield a 1:1 mixture of diastereomers of DHA (**105**) as white solid (412.6 mg, 81%). M.p. 157-158 °C (Lit.[392] 159 °C).

^{1}H-NMR (400 MHz, $CDCl_3$) δ 5.60 (s, 0.5 × H-5β), 5.39 (s, 0.5 × H-5α), 5.30 (d, *J* = 3.3 Hz, 0.5 × H-12β), 4.75 (d, *J* = 9.1 Hz, 0.5 × H-12α), 2.61 (m, H-11β), 2.37 (td, *J* = 14.2 Hz, 3.9 Hz, H-3a), 2.32 (m, H-11α), 2.03 (dd, *J* = 15.0 Hz, 3.1 Hz, H-3b), 1.95 – 1.18 (m, H-1, (H-2)$_2$, H-7, (H-8)$_2$, (H-9)$_2$, H-10), 1.43 (s, (H-15)$_3$), 0.95 (d, *J* = 6.4 Hz, (H-14)$_3$), 0.91 (d, *J* = 7.1 Hz, (H-13)$_3$). Mass spectrum (ESI+) *m/z:* 267 ([M-H_2O]$^+$).

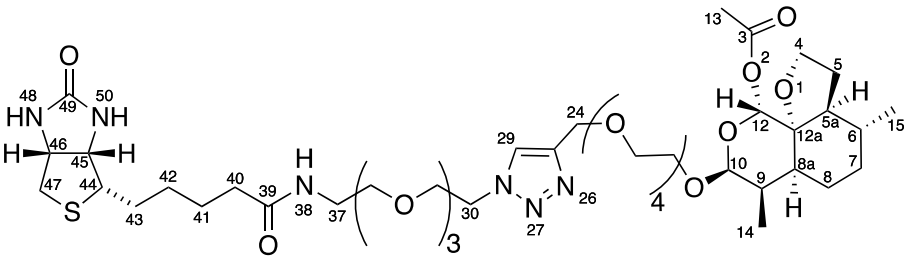
5.3.2.20 Acetylene-TEG-DHA[*] (107)



A solution of DHA (**105**; 50 mg, 176 µmol) and acetylene-TEG-OH (**76**; 122 mg, 528 µmol) in anhydrous DCM (500 µL) was stirred for 5 min at room temperature. Trichloroacetonitrile (20 µL, 194 µmol) and anhydrous tin dichloride (1.7 mg, 8.8 µmol) were added to the solution and the reaction mixture was kept stirring for 16 h. The reaction mixture was loaded onto a silica column equilibrated in 99:1 DCM:methanol. A gradient of 1-10% methanol:DCM was applied yielding two stereoisomers (C-12α and C-12β; ratio 1:1) of product **107** as mixture (total 87.6 mg, 38.0%).

^{1}H-NMR (400 MHz, $CDCl_3$) δ 5.41 (s, 0.5 × H-5β), 5.37 (s, 0.5 × H-5α), 4.82 (d, *J* = 3.5 Hz, 0.5 × H-12β), 4.73 (d, *J* = 9.3 Hz, 0.5 × H-12α), 4.20 (d, *J* = 2.3 Hz, $(H\text{-}28)_2$), 3.98 - 3.56 (m, $(H\text{-}16)_2$ - $(H\text{-}26)_2$), 2.61 (m, H-11β), 2.43 (t, *J* = 2.3 Hz, H-30), 2.36 (td, *J* = 14.0 Hz, 4.0 Hz, H-3a), 2.32 (m, H-11α), 2.03 (dd, *J* = 15.0 Hz, 3.1 Hz, H-3b), 1.95 – 1.18 (m, H-1, $(H\text{-}2)_2$, H-7, $(H\text{-}8)_2$, $(H\text{-}9)_2$, H-10), 1.42 (s, $(H\text{-}15)_3$), 0.94 (d, *J* = 6.4 Hz, $(H\text{-}14)_3$), 0.90 (d, *J* = 7.1 Hz, $(H\text{-}13)_3$). ^{13}C-NMR (100 MHz, $CDCl_3$) δ 102.1 (C-4), 91.2 (C-12), 87.9 (C-5), 81.2 (C-29), 80.4 (C-6), 74.7 (C-30), 70.9-67.8 (8C, C-16 – C-26), 58.4 (C-28), 52.6 (C-1), 44.5 (C-7), 37.5 (C-10), 36.5 (C-3), 34.7 (C-9), 30.9 (C-11), 26.2 (C-15), 24.8 (C-2), 24.5(C-8), 20.4 (C-14), 13.0 (C-13). Mass spectrum (ESI+) *m/z:* 521 ($[M+Na]^+$). (HRESI+) Found *m/z*: 521.2718, $C_{26}H_{42}O_9Na$ requires 521.2727.

5.3.2.21 Biotin-DHA-derivative[*] (109)

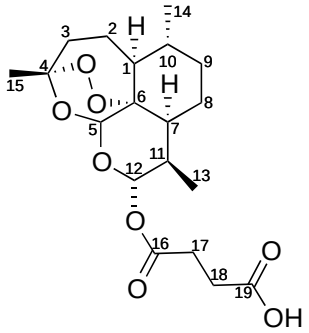


[*] denotes a new compound

A solution of acetylene-TEG-β-DHA (**107**; 4 mg, 8.03 μmol) and biotin-TEG-N_3 (**82**; 3.6 mg, 8.1 μmol) in anhydrous DMSO (500 μL) was stirred for 15 min at room temperature and added to copper iodide (0.08 mg; 0.4 μmol) and kept stirring for additional 24 h. The reaction mixture was diluted with water (2 mL) and loaded onto a disposable Strata-X cartridge (100 mg bed weight). After washing with 95:5 water:methanol (5 mL) the mixture was eluted with methanol (3 mL). The solvent was removed under a stream of nitrogen and the residue was purified by HPLC (semi-preparative C18 column, 4.7 mL/min, gradient from 5-70% acetonitrile:water over 30 min). The peak eluting after 21.7 min was collected, acetonitrile was removed under a stream of nitrogen and the water lyophilised to yield biotin-DHA-derivative (**109**) as light yellow oil (4.8 mg, 63.2%).

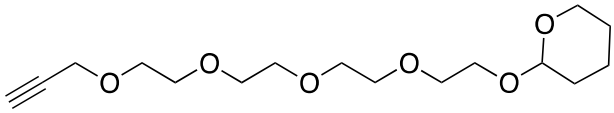
ν_{max} (Neat film) 3396 (m), 2875 (m), 2360 (m), 1751 (s), 1696 (s), 1650 (s), 1555 (m), 1457 (s), 1229 (s), 1090 (s), 1024 (m), 823 (s), 761 (s), 615 (m) cm^{-1}. ^{1}H-NMR (600 MHz, DMSO-d_6) δ 8.04 (s, H-29), 7.82 (t, *J* = 5.6 Hz, H-38), 6.41 (bs, H-48), 6.34 (bs, H-50), 6.07 (s, H-12), 4.67 (d, *J* = 4.2 Hz, H-10), 4.50 (s, $(H\text{-}24)_2$), 4.49 (t, *J* = 5.1 Hz, $(H\text{-}30)_2$), 4.29 (m, H-46), 4.11 (m, H-45), 4.03 (q, *J* = 7.9 Hz, H-4), 3.79 (m, $(H\text{-}31)_2$), 3.78 (m, H-16), 3.74 (q, *J* = 7.9 Hz, H-4), 3.54 (m, $(H\text{-}23)_2$), 3.52 (m, $(H\text{-}17)_2$), 3.45 - 3.56 (m, $(H\text{-}18)_2$ - $(H\text{-}22)_2$, $(H\text{-}32)_2$ - $(H\text{-}35)_2$), 3.39 (m, H-16), 3.36 (t, *J* = 6.0, $(H\text{-}36)_2$), 3.16 (m, $(H\text{-}37)_2$), 3.08 (m, H-44), 2.81 (dd, *J* = 12.5 Hz, 5.1 Hz, H-47β), 2.56 (d, *J* = 13.4 Hz, H-47α), 2.21 (m, H-9), 2.05 (t, *J* = 7.5, $(H\text{-}40)_2$), 2.07 (s, $(H\text{-}13)_3$), 1.86 (m, H-8), 1.85 (m, H-5), 1.77 (m, H-7), 1.75 (m, H-5), 1.69 (m, H-8), 1.59 (m, H-43), 1.52 (m, H-8a), 1.48 (m, $(H\text{-}41)_2$), 1.44 (m, H-43), 1.36 (m, H-6), 1.28 (m, $(H\text{-}42)_2$), 1.23 (m, H-5a), 0.86 (m, H-7), 0.86 (d, *J* = 6.4 Hz, $(H\text{-}14)_3$), 0.81 (d, *J* = 7.4 Hz, $(H\text{-}15)_3$). ^{13}C-NMR (150 MHz, DMSO-d_6) δ 172.0 (C-39), 169.0 (C-3), 162.7 (C-49), 143.7 (C-25), 124.2 (C-29), 101.1 (C-10), 87.5 (C-12), 79.6 (C-12a), 69.4 - 69.9 (C-18 – C-22 and C-32 – C-35), 69.3 (C-17, 69.1 (C-36), 69.0 (C-23), 68.9 (C-31), 67.5 (C-4), 66.7 (C-16), 63.5 (C-24), 61.0 (C-45), 59.2 (C-46), 55.3 (C-44), 55.0 (C-5a), 49.3 (C-30), 46.2 (C-8a), 39.8 (C-47), 38.4 (C-37), 35.2 (C-7), 35.0 (C-40), 32.8 (C-9), 30.2 (C-6), 28.2 (C-42), 28.0 (C-43), 27.1 (C-5), 25.2 (C-41), 24.4 (C-8), 21.2 (C-13), 20.3 (C-14), 12.2 (C-15). Mass spectrum (ESI+) *m/z:* 966 ($[M+Na]^+$). (HRESI+) Found *m/z*: 965.4871, $C_{44}H_{74}N_6O_{14}SNa$ requires 965.4881.

5.3.2.22 Artesunate (106)



Dihydroartemisinin (**105**; 40.0 mg, 140.7 μmol) was stirred in DCM (1 mL) for two minutes at room temperature. Freshly recrystallised ($CHCl_3$) succinic anhydride (22.6 mg, 225.8 μmol) and imidazole (8.6 mg, 126.3 μmol) were added to this solution and stirred for 1h. The reaction mixture was diluted with DCM (4 mL) and the pH of the reaction was adjusted to 5-6. The organic layer was washed with water, dried over magnesium sulfate and concentrated *in vacuo* to an oily mass. This residue was taken up in a small volume of petroleum ether:ethyl acetate (1:1) and loaded onto a silica column that had been equilibrated in 9:1 petroleum ether:ethyl acetate. The column was developed with a gradient of 10-50% ethyl acetate:petroleum ether yielding exclusively the α-artesunate (**106**) as a colourless solid (49.1 mg, 91%). M.p. 134-136 °C (Lit.[393] 134-137 °C). ^{1}H-NMR (400 MHz, $CDCl_3$) δ 5.63 (d, *J* = 9.8 Hz, H-12α), 5.54 (s, H-5), 2.74 - 2.63 (m, $(H\text{-}17)_2$ - $(H\text{-}18)_2$), 2.27 (m, H-11), 2.17 (ddd, *J* = 14.5 Hz, 13.4 Hz, 4.0 Hz, H-3a), 1.98 (m, H-3b), 1.80 (m, H-2a), 1.62 (m, H-8a), 1.59 (m, H-9a), 1.53 (m, H-7), 1.44 (m, H-8b), 1.41 (m, H-10), 1.31 (m, H-2b), 1.27 (s, $(H\text{-}15)_3$), 1.16 (m, H-1), 0.93 (m, H-9b), 0.88 (d, *J* = 6.5 Hz, $(H\text{-}14)_3$), 0.75 (d, *J* = 7.2 Hz, $(H\text{-}13)_3$).

5.3.2.23 Acetylene-TEG-pyran (116; 2-(3,6,9,12-tetraoxapentadec-14-ynyloxy)tetrahydro-2H-pyran)*

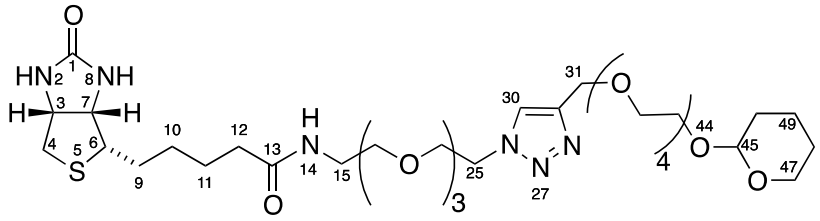


Pyridinium *p*-toluenesulfonate (2.2 mg, 8.6 μmol) in dry DCM (0.1 mL) was added to a solution of compound **76** (20.0 mg, 86.2 μmol) and 2,3-dihydropyran (12 μL, 130 μmol) in dry DCM (0.6 μL) and stirred for 5 h. The reaction mixture was diluted with DCM (5 mL) and washed with half saturated brine (2 × 3 mL). The aqueous layer was backwashed with diethylether (2 × 5 mL) and the organic layers were combined,

* denotes a new compound

dried over anhydrous magnesium sulfate and reduced to dryness in vacuo to give **116** as a yellow oil (12.3 mg, 45.1%). Mass spectrum (ESI+) *m/z*: 339 ($[M+Na]^+$), 317 ($[M+H]^+$). The oil was used immediately to make Biotin-TEG-triazole-TEG-pyran (**117**) and not further characterised.

5.3.2.24 Biotin-TEG-triazole-TEG-pyran (117)*

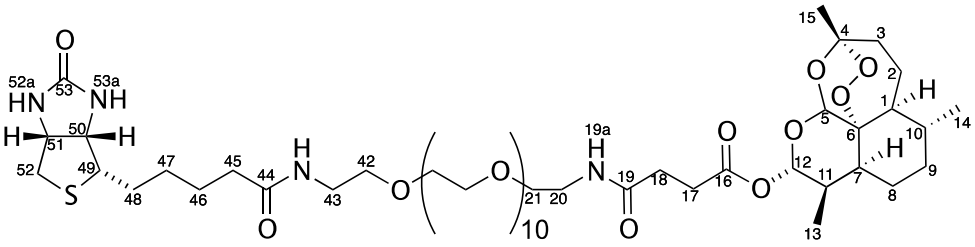


Copper(II) sulfate (0.1 mM; 57 µL, 5.7 µmol) was added to a solution of sodium ascorbate (2.3 mg, 11.4 µmol) in ethanol:water (1:1, 0.25 mL) and stirred until the initially formed brown precipitate turned yellow (5 min). To this, a solution of Biotin-TEG-azide (**82**; 25.3 mg, 57 µmol) and Tetrahydropyran-TEG-acetylene (**116**; 30.8 mg, 91.8 µmol) in ethanol:water (1:1, 1 mL) was added and stirred for 18 h. The reaction mixture was diluted with 95:5 water:methanol (2 mL) and loaded onto a disposable Strata-X cartridge (100 mg bed weight). After washing with 95:5 water:methanol (5 mL) the mixture was eluted with absolute methanol (3 mL). The solvent was removed under a stream of nitrogen and the residue was purified by HPLC (semi-preparative C18 column, 4.7 mL/min, gradient from 15-55% acetonitrile:water over 20 min). The peak eluting after 15.5 min was collected, acetonitrile was removed under a stream of nitrogen and the water lyophilised to give Biotin-TEG-triazole-TEG-pyran (**117**) as a white solid (7.5 mg, 25.9%). ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.78 (s, H-30), 6.75 (t, *J* = 5.4 Hz, H-14), 6.14 (s, H-8), 5.23 (s, H-2), 4.67 (s, $(H\text{-}31)_2$), 4.62 (t, *J* = 4.7 Hz, $(H\text{-}43)_2$), 4.55 (t, *J* = 5.2 Hz, $(H\text{-}25)_2$), 4.49 (m, H-3), 4.31 (m, H-7), 3.91 – 3.37 (m, $(H\text{-}15)_2$ – $(H\text{-}24)_2$ and $(H\text{-}33)_2$ – $(H\text{-}42)_2$), 3.13 (m, H-6), 2.89 (dd, *J* = 12.7 Hz, 4.8 Hz, H-4β), 2.71 (d, *J* = 12.8, H-4α), 2.20 (t, *J* = 7.5 Hz, $(H\text{-}12)_2$), 1.96 – 1.37 (m, $(H\text{-}9)_2$ –($H\text{-}11)_2$, H-45 and $(H\text{-}47)_2$ – $(H\text{-}50)_2$). ^{13}C-NMR (100 MHz, $CDCl_3$) δ 173.4 (C-13), 163.6 (C-1), 145.0 (C-29), 124.0 (C-30), 99.0 (C-45), 70.9 – 69.5 (C-16 – C-24 and C-33 – C-42), 66.7 (C-43), 64.6 (C-31), 62.3 (C-47), 61.8 (C-7), 60.2 (C-3), 55.5 (C-6), 50.2 (C-25), 40.6 (C-4), 39.2 (C-15), 35.9 (C-12), 30.6 (C-50), 28.14 (C-10), 28.07 (C-9), 25.6 (C-11), 25.4 (C-48), 19.6 (C-49). Mass spectrum (ESI+)

* denotes a new compound

m/z: 761 ([M+H]$^+$. (HRESI+) Found *m/z*: 783.3928, $C_{34}H_{60}N_6O_{11}SNa$ requires 783.3933.

5.3.2.25 Biotin-α-artesunate (123)*



To a cooled (0 °C) mixture of α-artesunate (21.8 mg, 56.7 μmol), EDC (21.7 mg, 113.3 μmol), and HOBt (15.3 mg, 113.3 μmol) in anhydrous DMF (0.50 mL) was added the amine (**84**; 29.1 mg, 37.8 μmol) in cold anhydrous DMF (0.25 mL) dropwise. The reaction mixture was stirred at 0 °C for 0.5 h and was then allowed to warm to ambient temperature. The mixture was stirred at room temperature for 18 h and concentrated under vacuum. The solution was then diluted with water (5 mL) and passed through a disposable C18 cartridge (500 mg bed weight). The column was washed with water (10 mL) to remove the DMF. The product was eluted with absolute acetonitrile and reduced to dryness under a stream of nitrogen. The residue was taken up in 5:95 acetonitrile:water and separated on HPLC (preparative C18 column, 9.99 mL/min, gradient from 15-70% acetonitrile:water over 30 min). The peak eluting after 23.8 min was collected, acetonitrile was removed under a stream of nitrogen and the water lyophilised to give α-artesunate-PEG-biotin (**123**) as a white solid (16.7mg, 38.9%).

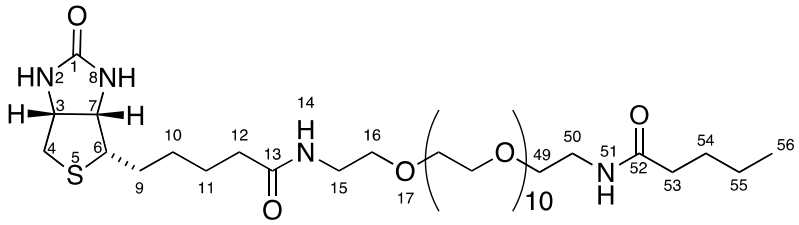
^{1}H-NMR (600 MHz, DMSO-d_6) δ 7.96 (t, *J* = 5.6 Hz, H-19a), 7.83 (t, *J* = 5.6 Hz, H-43a), 6.41 (s, H-53a), 6.35 (s, H-52a), 5.63 (d, *J* = 9.8 Hz, H-12α), 5.54 (s, H-5), 4.29 (m, H-51), 4.11 (m, H-50), 3.75 – 3.42 (m, $(H\text{-}22)_2$ – $(H\text{-}41)_2$), 3.38 (t, *J* = 6.1 Hz, $(H\text{-}21)_2$), 3.37 (t, *J* = 6.0 Hz, H-42), 3.17 (m, $(H\text{-}20)_2$ and $(H\text{-}43)_2$), 3.08 (m, H-49), 2.80 (dd, *J* = 12.5 Hz, 5.1 Hz, H-52β), 2.57 (m, H-17), 2.56 (d, *J* = 13.4 Hz, H-52α), 2.38 (m, $(H\text{-}18)_2$), 2.27 (m, H-11), 2.17 (ddd, *J* = 14.5 Hz, 13.4 Hz, 4.0 Hz, H-3a), 2.05 (dd, *J* = 7.5 Hz, 7.5 Hz, $(H\text{-}45)_2$), 1.98 (m, H-3b), 1.80 (m, H-2a), 1.62 (m, H-8a), 1.59 (m, H-9a), 1.53 (m, H-7), 1.48 (m, $(H\text{-}46)_2$), 1.45 (m, $(H\text{-}48)_2$), 1.44 (m, H-8b), 1.41 (m, H-10), 1.31 (m, H-2b), 1.28 (m, $(H\text{-}47)_2$), 1.27 (s, $(H\text{-}15)_3$), 1.16 (m, H-1), 0.93 (m, H-9b), 0.88 (d, *J* = 6.4

* denotes a new compound

Hz, (H-14)$_3$), 0.75 (d, *J* = 7.2 Hz, (H-13)$_3$). ^{13}C-NMR (150 MHz, DMSO-d_6) δ 172.1 (C-44), 171.4 (C-16), 170.7 (C-19), 162.8 (C-53), 103.6 (C-4), 91.6 (C-12), 90.6 (C-5), 79.9 (C-6), 69.2 (C-42), 69.1 (C-21), 61.1 (C-50), 59.2 (C-51), 55.4 (C-49), 51.1 (C-1), 44.6 (C-7), 39.8 (C-52), 38.6 (C-20), 38.4 (C-43), 36.0 (C-10), 35.9 (C-3), 35.1 (C-45), 33.6 (C-9), 31.7 (C-11), 29.5 (C-18), 28.9 (C-17), 28.2 (C-47), 28.0 (C-48), 25.5 (C-15), 25.2 (C-46), 24.2 (C-2), 21.0 (C-8), 20.1 (C-14), 11.7 (C-13). Mass spectrum (ESI+) *m/z:* 591 ([M+2Na]$^{2+}$). (HRESI+) Found *m/z*: 1159.5920, $C_{53}H_{92}N_4O_{20}SNa$ requires 1159.5918.

The purity of the compound was verified by reinjecting an aliquot (30 µL; 0.88 mM in 30% acetonitrile/water) onto an analytical C18 RP-HPLC column (1 mL/min, 30-80% acetonitrile over 30 min). The trace contained exclusively one peak, which eluted after 18.5 min and showed an absorbance maximum at 190 nm.

5.3.2.26 Biotin-valeramide (124)*



To a cooled (0 °C) mixture of valeric acid (8.55 µL, 77.9 µmol), EDC (22.4 mg, 116.8 µmol), and HOBt (15.8 mg, 116.8 µmol) in anhydrous DMF (0.7 mL) was added the amine (**84**; 30.0 mg, 38.9 µmol) dissolved in cold anhydrous DMF (0.3 mL) dropwise. The reaction mixture was stirred at 0 °C for 0.5 h and was then allowed to warm to ambient temperature. The mixture was stirred at room temperature for 18 h, and concentrated under vacuum. The solution was then diluted with water (5 mL) and passed through a disposable C18 cartridge (500 mg bed weight). The column was washed with water (10 mL) to remove the DMF and the reaction mixture eluted with absolute acetonitrile. The mixture was reduced to dryness under a stream of nitrogen and the residue taken up in 5:95 acetonitrile:water and separated on HPLC (semi-preparative C18 column, 4.7 mL/min, gradient from 15-70% acetonitrile:water over 30 min). The peak eluting after 16.2 min was collected, acetonitrile was removed under a stream of nitrogen and the water lyophilised to give valeric acid-PEG-biotin-amide (**124**) as white solid (8.6 mg, 28.5%).

* denotes a new compound

^{1}H-NMR (400 MHz, methanol-d_4) δ 4.52 (m, H-3), 4.34 (m, H-7), 3.80 – 3.46 (m, (H-16 – H-49)$_2$), 2.93 (dd, *J* = 15.8 Hz, 5.2 Hz, H-4β), 2.73 (d, *J* = 12.7 Hz, H-4α), 2.25 (t, *J* = 7.5 Hz, (H-12)$_2$), 2.22 (t, *J* = 7.6 Hz, (H-50)$_2$), 1.76 – 1.27 (m, (H-9)$_2$ – (H-11)$_2$ and (H-54)$_2$ – (H-55)$_2$), 1.20 (t, *J* = 7.0 Hz, (H-53)$_2$), 0.96 (t, *J* = 7.3 Hz, (H-56)$_3$). ^{13}C-NMR (150 MHz, methanol-d_4) δ 68.6 – 67.8 (C-16 – C-49), 60.5 (C-7), 58.8 (C-3), 54.2 (C-6), 38.2 (C-4), 37.42 (C-15), 37.38 (C-50), 33.9 (C-12), 33.8 (C-53), 26.9 (C-11), 26.6 (C-10), 26.3 (C-54), 24.0 (C-9), 20.5 (C-55), 11.3 (C-56). Mass spectrum (ESI+) *m/z:* 440 ([M+H+Na]$^{2+}$), 429 ([M+2H]$^{2+}$). (HRESI+) Found *m/z*: 877.4830, $C_{39}H_{74}N_4O_{14}SNa$ requires 877.4815.

5.4 Phage Display

Biosafety approval was obtained from the Macquarie University Biosafety Committee (approval number NLRD 5201001557).

5.4.1 Overview

The display cloning protocols presented below were adapted from those described in the Novagen T7Select Manual[394].

5.4.2 Materials

Sodium chloride, potassium dihydrogen phosphate, IPTG, DNA molecular weight markers and carbenicillin were obtained from Sigma-Aldrich (USA). Tryptone, yeast extract, agar and polystyrene Petri dishes were obtained from Bacto Laboratories (Australia). Glucose, agarose, super-fine resolution (SFR) agarose and Tris were purchased from AMRESCO (USA). Acetic acid, glycerol, ammonium chloride, disodium hydrogen phosphate and EDTA disodium salt were obtained from BDH (Germany). T7Select10-3 human disease cDNA libraries and *E. coli* strain BLT5615 was obtained from Novagen Inc. (USA). T7Select1-1b human brain cDNA library was a kind gift from Prof. David Austin (Department of Chemistry, Yale University). Nucleotides (dNTPs) were obtained from Bioline (UK). Oligonucleotides (primers) were obtained from Sigma-Genosys (Australia). *Taq* DNA polymerase and QIAquick PCR purification kits were obtained from QIAGEN (USA). *Hin*fI restriction endonuclease and NEB buffer 2 were obtained from Promega Corp. (USA). Electrophoresis grade agarose was obtained from American Bioanalytical (USA). Nuclease-free water, 1 M magnesium chloride and 20% SDS were obtained from Ambion (USA). Reacti-bind HBC Neutravidin 8-well strip plates were obtained from Pierce (USA). Disposable plastic syringes were obtained from Terumo (Japan). Polystyrene 96-well microtitre plates, flexible poly(vinyl chloride) 96-well assay plates and conical 250 mL centrifuge bottles were obtained from Corning (USA).

Table 26: Preparation of reagents and media used in display cloning

Reagent	Ingredients	Instructions
20xM9	20 g NH_4Cl 60 g KH_2PO_4 120 g $Na_2HPO_4.12H_2O$ Water to make 1 L of solution	- Adjust pH to 7.4 - Autoclave 20 min / 121 °C
20% Glucose	200 g glucose Water to make 1 L of solution	- Autoclave 20 min / 121 °C
5% Carbenicillin	1 g carbenicillin disodium Water to make 20 mL of solution	- Sterile filter (0.22 µm) - Store at -20 °C
50xTAE	242 g Tris 57.1 mL acetic acid 18.6 g EDTA disodium salt Water to make 1 L of solution	- Adjust pH to 8.0 - Autoclave 20 min / 121 °C
M9TB	9.28 g tryptone 4.64 g NaCl 50 mL 20x M9 1 mL $MgCl_2$ (1M) Water to make 1 L of solution	- Autoclave 20 min / 121 °C - Cool to room temperature - Add 20 mL 20% glucose - Add 1 mL 5% carbenicillin
2xYT	17 g tryptone 10 g yeast extract 5 g NaCl Water to make 1 L of solution	- Adjust pH to 7.4 - Autoclave 20 min / 121 °C
LB Agar	10 g tryptone 5 g yeast extract 10 g NaCl 15 g agar Water to make 1 L of solution	- Adjust pH to 7.4 - Autoclave 20 min / 121 °C - Cool to 55 °C - Pour plates (~100) - Store plates at 4 °C
LB Agarose	10 g tryptone 5.2 g yeast extract 5.2 g NaCl 6 g agarose Water to make 1 L of solution	- Adjust pH to 7.4 - Autoclave 20 min / 121 °C - Cool to 55 °C - Store in 25 mL lots at 4 °C
PCR Master Mix	10 µL 'UP' Primer (10 µM) 10 µL 'DOWN' Primer (10 µM) 20 µL dNTPs (10 mM each dNTP) 100 µL Taq buffer (10x) 835 µL nuclease free water	- Store at –20 °C - Defrost and add 5 µL TaqPolymerase before use
DNA Finderprinting Mix	10 µL *Hin*fI enzyme (1000 U/mL) 50 µL NEB Buffer 2 (10x) 240 µL nuclease free	- Can be stored at 4 °C for short periods
DNA Loading Buffer (6x)	40 g sucrose 100 mg bromophenol blue Water to make 100 mL	- Sterile filter (0.22 µm) - Store at 4 °C

5.4.3 Equipment

Bacterial cultures were incubated in a heated orbital shaker (Thermoline Scientific, Australia). Optical densities were recorded in 1 cm polystyrene semi-micro cuvettes (Sarstedt, Germany) using a BioRad SmartSpec Plus UV spectrophotometer using a 600nm filter (BioRad, USA). Solutions were centrifuged with a 6K15 refrigerated centrifuge (Sigma, Germany). DNA was amplified with a C1000 Thermal Cycler (Bio-Rad, USA). DNA sequencing was performed by the Macquarie University DNA Analysis Facility using a *3130xl Genetic Analyzer* (Applied Biosystems, USA). Agarose gel electrophoresis was performed using a Mini-Sub Cell GT system (BioRad, USA) and gels were visualised via Gel-Red stain with a G:Box Chemi transilluminator (ethidium bromide filter) using GeneSnap digital imaging software (SynGene, UK). Water was purified using a Milli-Q Ultrapure Water Purification System (Millipore, USA). Linear and non-linear least squares regression analysis was performed using GraphPad Prism 4.0 for Macintosh (GraphPad Software, USA).

5.4.4 Preparation of bacterial cultures

Stocks of *E. coli* strains BLT5615 were stored at –80 °C in 10% glycerol. An initial culture was prepared by streaking a small quantity of this frozen stock onto an LB agar plate and incubating the plate at 37 °C for 16 h (BLT5615) or 24 h (B5615). The plate carrying BLT5615 was stored at 4 °C, whereas the slowly growing B5615 was kept at room temperature. Each plate was used to inoculate subsequent cultures for up to three weeks.

A saturated overnight culture of BLT5615 or B5615 was prepared by inoculating M9TB (20 mL) with a single bacterial colony from an LB agar plate and then incubating at 37 °C for 16 h with gentle swirling (120 - 150 rpm).

A fresh culture of BLT5615 or B5615, ready for infection by T7 bacteriophage, was prepared by inoculating M9TB (100 mL) with saturated overnight culture (5 mL) and incubating at 37 °C with vigorous shaking until an OD_{600} of 0.4 was reached (1.5-3 h). IPTG (24%; 100 µL) was added and incubation continued for a further 30 min (BLT5615) or 60 min (B5615). The culture was then stored on slushy ice (for up to 24 h) until required.

5.4.5 Growth of T7 lysates

IPTG-treated cells (100 mL, BLT5615 or B5615) were infected with a T7Select cDNA library (1 µL) and incubated at 37 °C with vigorous shaking until lysis had occurred (1-2 h), as indicated by a marked decrease in OD_{600}. Immediately following lysis, the lysate was centrifuged at 4700 rpm for 10 min at 4 °C to precipitate cellular debris and the supernatant was decanted into a clean tube containing Tween-20 (1%; 1 mL). The clarified lysate containing 0.01% Tween-20 was stored on slushy ice until required.

5.4.6 Preparation of natural product derivatised microtitre plates

5.4.6.1 Stock solutions of biotinylated natural products and controls

1 µmol/mL solutions of all control probes or natural product probes were made up in DMSO and stored at -80 °C. Dilutions (1:100) in PBS resulted in 10 nmol/mL solutions and these were stored at -20 °C for up to two weeks.

5.4.6.2 Biotinylated natural products on neutravidin-coated PS plates

All microtitre plate wells were pre-incubated with PBS (250 µL) for 1 h at room temperature before use. The wells were emptied and 100 µL probe solutions (10 nmol/mL) were applied for 2 h at room temperature or for 16 h at 4 °C. The supernatant was removed, each well washed with PBS (3 × 250 µL) and immediately used for affinity selection.

5.4.7 Affinity selections

Clarified T7 phage lysate (200 µL) was added to one well of a neutravidin-coated PS plate that had been derivatised with a biotinylated control compound, and was left to incubate for 1 h at room temperature. The lysate was then transferred to a second well of the plate that had been derivatised with the biotinylated target molecule, and was left to incubate for 3 h at room temperature. The well was washed with PWB (3 × 250 µL) and incubated with SDS (1%; 100 µL) for 30 min at room temperature. Finally, eluate was diluted with 2xYT (1:10; 900 µL) and stored at 4 °C overnight, during which time a portion of the SDS precipitated. The following day, an aliquot of the eluate (1:10 in 2xYT; 20 µL) was removed, taking care not to disturb any precipitated SDS, and added to fresh IPTG-treated BLT5615 or B5615 *E. coli* cells (20 mL) for the next round of selection. This procedure was repeated until 7-12 rounds of selection had been completed. The stringency of the washing step was increased

with each successive round of selection, from 3 × 250 µL PWB over 10 s in Round 1 to 5 × 250 µL PWB over 2 min in Rounds 7-12.

5.4.8 Titring

Standard, round, LB agar plates were prewarmed to 37 °C. LB agarose (5 mL) was completely molten in a microwave oven and allowed to cool to 50 °C. IPTG-treated BLT 5615 cells (250 µL, OD_{600} = 0.8-1) and IPTG (24%; 5 µL) were added to the cooled agarose and the mixture was poured onto one LB agar plate. To allow for the agarose to completely set, the plate was kept uncovered at room temperature for 30-45 min. The phage eluate retained from each round of selection was serially diluted with 2xYT medium from 10^{-1} to 10^{-10} in a flexible 96-well assay plate. A small aliquot (2 µL) of each dilution from each round of selection was dropped onto the surface of the solidified agarose using a multi-channel micropipette (8 × 5 array per plate). The uncovered plate was left to stand at room temperature until the drops had adsorbed entirely into the agarose. Consecutively, each plate was incubated for 2-3 hours at 37 °C until plaques were clearly visible against the lawn of bacteria. The phage titre was calculated from that particular phage dilution of each round of selection, which contained a countable number (5-50) of plaques.

5.4.9 Picking plaques

Serial dilutions (10^{-1} to 10^{-7}) with 2xYT were prepared from amplified phage lysate from the final round of selection. LB agarose (5 mL) was completely molten in a microwave oven and allowed to cool to 50 °C. IPTG-treated BLT 5615 cells (250 µL, OD_{600} = 0.8-1), IPTG (24%; 5 µL) and an aliquot (50 µL) of the 10^{-7} dilution were added to the cooled agarose and the mixture was poured onto one LB agar plate. After allowing for the agarose to completely settle, the plate was incubated at 37 °C until plaques were clearly visible against the lawn of bacteria (2-4 h). Individual plaques (24) were collected by stabbing the centre of each plaque with a 10 µL micropipette tip and transferring the tip to IPTG-treated BLT5615 cells (OD_{600} = 0.6–0.8; 100 µL) in a 96-well microtitre plate. The tips were removed and the plate was incubated until complete lysis of bacterial cells occurred in each well (1-2 h). The plate was centrifuged at 4300 rpm for 10 min at 4 °C. An aliquot (40 µL) of the supernatant was transferred into a clean 96-well microtitre plate containing 80% glycerol (10 µL per well) and stored at -80 °C until required.

5.4.10 Amplification, sequencing and fingerprinting of cDNA inserts

A solution of phage lysate (0.5 µL) and PCR master mix (19.5 µL, including Taq Polymerase) was prepared and subjected to 20 rounds of thermocycling using the protocol shown in Table 27. An aliquot of the amplified DNA solution (2 µL) was then incubated with the DNA fingerprinting mix (4 µL) at 37 °C for 1h.

Table 27: Standard thermocycler program for PCR of cDNA inserts

Cycles	Temperature	Time
1	94 °C	150 s
20	94 °C 55 °C 72 °C	45 s 60 s 30 s
1	72 °C	10 min
1	4 °C	∞

5.4.11 Gel electrophoresis

Electrophoresis-grade agarose (0.6 g) was suspended in 1 × TAE buffer (40 mL) and the suspension was boiled in a microwave oven until the agarose had dissolved completely. The 1.5% solution was poured into a casting tray (10 × 7 cm) containing two 15 well combs, and allowed to set for 30-45 min at room temperature. Once the gel had solidified, it was transferred to a gel tank, flooded with 1 × TAE, and the combs were removed. For gel electrophoresis of PCR products of single plaques picked from the final round of selection, the agarose concentration was increased to 2%. All digested fingerprinting samples were run in gels made up of super-fine resolution agarose (3%).

Each amplified cDNA insert of digested fingerprinting sample (5 µL) was mixed with 6 × DNA loading buffer (1 µL) and loaded onto the gel with a micropipette. After all samples had been loaded, the gel was run at 80V until the bromophenol blue dye had migrated approximately half way down each half of the gel (25-30 min). The gel was then removed from the tank and submerged in a Gel-Red® post-staining solution (3.3 ×) for 60 min. After de-staining in deionised water (10 min), the gel was visualised using an a:Box Chemi UV transilluminator.

5.4.12 DNA sequencing

An aliquot of PCR-amplified DNA (10 μL) was purified using a QIAquick PCR purification kit following the manufacturer's instructions, providing 30 μL solution containing the purified DNA. Of this, an aliquot (8 μL) was combined with one PCR primer (1 μM; 4 μL, 4 pmol) and the resulting solution was submitted for DNA sequencing.

5.4.13 Target validation (binding studies) - non-specific elution

The microtitre plates were preconditioned with PBS (250 μL) for 1 h at room temperature before use. A single plaque was reamplified in *E.coli* BLT5615 and the phage lysate clarified (see 5.4.5). Three wells of a microtitre plate were derivatised with a control compound and three others with the target molecule (see 5.4.6.2). An aliquot (100 μL) of clarified phage lysate was pipetted into each and the wells were left to incubate for 2 h at room temperature. The lysates were then aspirated and the wells were washed with PWB (10 × 250 μL; 4 °C). Any phage particles remaining in the well were eluted with SDS (1%; 100 μL) over 20min. Serial dilutions were made from the eluates with 2xYT and titred (see 5.4.8).

.

Chapter 6

Conclusions and Future Directions

The abject failure of combinatorial chemistry and more stringent mode of action requirements mean that modern drug discovery currently faces major bottlenecks in delivering new chemical entities (NCE) and in getting new drugs approved for human use. What is needed are more methods to rapidly identify new lead compounds and to elucidate their molecular modes of action. This thesis attempts to address this through the continuing development of phage display technologies for the elucidation of natural product targets. As all natural products have some function, phage display can quickly link this function to a gene if the natural product binds to a protein. This rapid correlation of small molecule, protein and gene can serve firstly to identify NCE's, druggable targets and indicate a path forward for mode of action studies.

In **Chapter One**, modern methods used to find protein binding partners for small molecules are reviewed and "reverse chemical proteomics" outlined as a tool to rapidly deliver information on protein targets that has many advantages over other technologies. The rationale for implementing natural products in this method are briefly discussed.

The overall importance of natural products in the drug development process are further highlighted in **Chapter Two**, which presents the isolation of marine natural products from sponges using a traditional bioassay-guided fractionation approach. The application of two primary assays to determine antibacterial and herbicidal activity of a small set of local sponges resulted in the isolation and structure elucidation of 9 alkaloids from the sponge *Pseudoceratina purpurea*. Two of the compounds were new to science. These compounds exhibited moderate antimicrobial activity. To elucidate if these new compounds also display other biological activities they could be biotinylated and employed in reverse chemical proteomics. This would link the results from Chapter Two directly to the achievements presented in Chapters Three and Four.

In **Chapter Three** we discuss the preparation of long, biotinylated polyethyleneglycol linkers with a range of terminal functionalities and highlight the importance of linker length when using small molecules to link biomolecules to a solid surface of some type. Chapter Three also presents the biotinylation of three structurally diverse, rare and bioactive small molecules, carried on a milligram scale, giving rise to three probes that can be employed in reverse chemical proteomics. A total of 16 new compounds were synthesized in this section.

In **Chapter Four** the application of the T7 phage display is described as a tool to identify cellular targets for biologically active natural products. To validate the system, we initially immobilised a biotinylated FK506 on a neutravidin-coated microtitre plate and the resulting affinity support was used to probe a T7 phage-displayed human brain cDNA library. The biopanning rescued two known receptors for FK506 (FKBP1a and FKBP1b) over seven rounds of selection from a plethora of other clones present in the original library.

We then applied the same assay to isolate the most avid protein binding partners for three biologically and structurally diverse natural products (manzamine, daptomycin and artesunate). For the depsipeptide antibiotic daptomycin, we isolated ribosomal protein S19 (RPS19) from four human libraries as multiple clones, providing good evidence that this is a specific target of daptomycin in humans. In the case of artesunate, the BCL-2 associated agonist of cell death (BAD) was isolated from five different libraries, again as multiple clones. BAD did not always dominate the sublibraries created in the final round of selection but all relevant clones displayed substantial parts or the coding sequence of the BAD protein on the phage surface. For both the artesunate and daptomycin probes, a non-specific on-phage binding assay provided further evidence to support the RPS19 and BAD respectively as being real targets for these small molecules. The results from the biopanning experiments with manzamine did not result in any one cellular target, despite doubling the number of biopanning rounds. These results indicate that either the protein binding partner is not present in the phage library, is poorly expressed or misfolded, was lost in the early rounds of biopanning, the natural product was rendered inactive by derivatisation, or that manzamine binds to other biomolecules (such as DNA or membranes) to achieve its biological activity. Regardless, the rapid convergence of the libraries of 2/4 natural products shows the utility of this method for the isolation of novel binding partners for small molecules. Naturally this system will not work for every compound but when it works, spectacular results are possible that can lead to rapid advances in drug development.

To address the first three reasons no protein was isolated for manzamine, we can use the biotinylated probe in a forward chemical proteomics approach to pull down all the proteins in a cell and to analyse these by MS. Ideally this could be done with the control probe in a stabile isotope labeling technique (e.g. ICAT[395] or SILAC[396])

experiments to quantitatively determine proteins that are enriched using the manzamine probe over the control probe. To address the later possibilities, we could attach the linker to another part of the molecule, such as the alkenes (e.g. by cross metathesis reaction with Grubb's catalyst as performed by Winkler *et al*.[274]) or use another techniques such as yeast display which may produce better folding results and with post-translational modifications.

Further *in vitro* work is required for validating the interactions of artesunate and daptomycin with BAD and RPS19 employing methods such as surface plasmon resonance (SPR; BIACore), differential scanning calorimetry (DSC; MicroCal) or microscale thermophoresis (MST; NanoTemper). Given that both natural products are available as biotinylated probes, streptavidin coated biosensor chips for the BIAcore can easily be utilised to immobilise the natural product on the gold surface. Concentration dependent binding studies with the isolated and purified protein would allow the *k*on/*k*off values to be determined for each interaction.

Initial attempts in use SPR or MST using the pure phage expressing the target proteins was not successful due to the low concentrations of phage (10–13 M) achievable in solution and the non-specific binding caused by the phage and cellular components of the host (*E. coli*). There are established protocols for cloning cDNA inserts from the phage into for example BL21 *E. coli* as an expression vector[397], so the next steps would be to clone and overexpress BAD and RPS19 and then apply these to the analytical techniques mentioned or even look at the interaction between protein and underivatised natural products by NMR spectroscopy (transfer-ROESY)[398]. Although quantitative measurements of the binding constants between the probe and its target are essential for further discussion of the specificity of the interaction, this next step is beyond the scope of this thesis but must be pursued to determine the true importance of the interactions discovered here.

Chapter 7

Bibliography

(1) Carlson, E. E. *ACS Chem Biol* **2010**, *5*, 639.
(2) Rouhi, A. M. *Chem Eng News* **2003**, *81*, 77.
(3) Cragg, G.; Newman, D. *Phytochem Rev* **2009**, *8*, 313.
(4) Newman, D. J.; Cragg, G. M. *J Nat Prod* **2007**, *70*, 461.
(5) Newman, D. J.; Cragg, G. M. *J Nat Prod* **2012**, *75*, 311.
(6) Azad, M. A.; Wright, G. D. *Bioorg Med Chem* **2012**, *20*, 1929.
(7) Wittgenstein, L. *Ann Naturphil* **1921**, *14*, 185.
(8) Schreiber, S. L. *Bioorg Med Chem* **1998**, *6*, 1127.
(9) Lehar, J.; Stockwell, B. R.; Giaever, G.; Nislow, C. *Nat Chem Biol* **2008**, *4*, 674.
(10) Yang, W. S.; Stockwell, B. R. *Chem Biol* **2008**, *15*, 234.
(11) Omura, S.; Fujimoto, T.; Otoguro, K.; Matsuzaki, K.; Moriguchi, R.; Tanaka, H.; Sasaki, Y. *J Antibiot* **1991**, *44*, 113.
(12) Pagano, M.; Tam, S. W.; Theodoras, A. M.; Beer-Romero, P.; Del Sal, G.; Chau, V.; Yew, P. R.; Draetta, G. F.; Rolfe, M. *Science* **1995**, *269*, 682.
(13) Kim, K. B.; Crews, C. M. *J Med Chem* **2008**, *51*, 2600.
(14) Gangadhar, N. M.; Firestein, S. J.; Stockwell, B. R. *Mol Cell Neurosci* **2008**, *38*, 518.
(15) Karuso, P. In *Comprehensive Natural Products Ii: Chemistry and Biology*; Mander, L., Lui, H. W., Eds.; Elsevier: Oxford, 2010, p 513.
(16) Rix, U.; Superti-Furga, G. *Nat Chem Biol* **2009**, *5*, 616.
(17) Keiser, M. J.; Irwin, J. J.; Shoichet, B. K. *Biochemistry* **2010**, *49*, 10267.
(18) Raju, T. N. K. *Lancet* **2000**, *355*, 1022.
(19) Barglow, K. T.; Cravatt, B. F. *Nat Methods* **2007**, *4*, 822.
(20) Sakamoto, S.; Hatakeyama, M.; Ito, T.; Handa, H. *Bioorg Med Chem* **2012**, *20*, 1990.
(21) Sadaghiani, A. M.; Verhelst, S. H. L.; Bogyo, M. *Curr Opin Chem Biol* **2007**, *11*, 20.
(22) de Guzman, F. S.; Copp, B. R.; Mayne, C. L.; Concepcion, G. P.; Mangalindan, G. C.; Barrows, L. R.; Ireland, C. M. *J Org Chem* **1998**, *63*, 8042.
(23) Margarucci, L.; Monti, M. C.; Tosco, A.; Riccio, R.; Casapullo, A. *Angew Chem, Int Ed Engl* **2010**, *49*, 3960.
(24) Busserolles, J.; Paya, M.; D'Auria, M. V.; Gomez-Paloma, L.; Alcaraz, M. J. *Biochem Pharmacol* **2005**, *69*, 1433.
(25) Monti, M. C.; Chini, M. G.; Margarucci, L.; Tosco, A.; Riccio, R.; Bifulco, G.; Casapullo, A. *J Mol Recognit* **2009**, *22*, 530.
(26) Margarucci, L.; Monti, M. C.; Fontanella, B.; Riccio, R.; Casapullo, A. *Mol BioSyst* **2011**, *7*, 480.
(27) Pearse, B. M. *Proc Natl Acad Sci U S A* **1976**, *73*, 1255.
(28) Chu, J. J. H.; Ng, M. L. *J Virol* **2004**, *78*, 10543.
(29) Blanchard, E.; Belouzard, S.; Goueslain, L.; Wakita, T.; Dubuisson, J.; Wychowski, C.; Rouille, Y. *J Virol* **2006**, *80*, 6964.
(30) Badis, G.; Chan, E. T.; van Bakel, H.; Pena-Castillo, L.; Tillo, D.; Tsui, K.; Carlson, C. D.; Gossett, A. J.; Hasinoff, M. J.; Warren, C. L.; Gebbia, M.; Talukder, S.; Yang, A.; Mnaimneh, S.; Terterov, D.; Coburn, D.; Li Yeo, A.; Yeo, Z. X.; Clarke, N. D.; Lieb, J. D.; Ansari, A. Z.; Nislow, C.; Hughes, T. R. *Mol Cell* **2008**, *32*, 878.
(31) Bantscheff, M.; Eberhard, D.; Abraham, Y.; Bastuck, S.; Boesche, M.; Hobson, S.; Mathieson, T.; Perrin, J.; Raida, M.; Rau, C.; Reader, V.; Sweetman, G.; Bauer, A.; Bouwmeester, T.; Hopf, C.; Kruse, U.; Neubauer, G.; Ramsden, N.; Rick, J.; Kuster, B.; Drewes, G. *Nat Biotech* **2007**, *25*, 1035.
(32) Lomenick, B.; Hao, R.; Jonai, N.; Chin, R. M.; Aghajan, M.; Warburton, S.; Wang, J.; Wu, R. P.; Gomez, F.; Loo, J. A.; Wohlschlegel, J. A.; Vondriska, T. M.; Pelletier, J.; Herschman, H. R.; Clardy, J.; Clarke, C. F.; Huang, J. *Proc Natl Acad Sci U S A* **2009**, *106*, 21984.
(33) Lomenick, B.; Olsen, R. W.; Huang, J. *ACS Chem Biol* **2011**, *6*, 34.
(34) Makowski, L.; Rodi, D. J.; Mandava, S.; Minh, D. D. L.; Gore, D. B.; Fischetti, R. F. *J Mol Biol* **2008**, *375*, 529.
(35) Henzler-Wildman, K.; Kern, D. *Nature* **2007**, *450*, 964.
(36) Piggott, A. M.; Karuso, P. *Comb Chem High Throughput Screening* **2004**, *7*, 607.
(37) Smith, G. S. *Science* **1985**, *228*, 1315.
(38) Sergeeva, A.; Kolonin, M. G.; Molldrem, J. J.; Pasqualini, R.; Arap, W. *Adv Drug Delivery Rev* **2006**, *58*, 1622.
(39) Buchholz, C. J.; Duerner, L. J.; Funke, S.; Schneider, I. C. *Comb Chem High Throughput Screening* **2008**, *11*, 99.
(40) Urban, J. H.; Merten, C. A. *ACS Chem Biol* **2010**, *6*, 61.
(41) Lee, S. Y.; Choi, J. H.; Xu, Z. *Trends Biotechnol* **2003**, *21*, 45.

(42) Ho, M.; Pastan, I. *Methods Mol Biol* **2009**, *562*, 99.
(43) Fields, S.; Song, O.-k. *Nature* **1989**, *340*, 245.
(44) Licitra, E. J.; Liu, J. O. *Proc Natl Acad Sci U S A* **1996**, *93*, 12817.
(45) Suwwan de Felipe, K.; Carter, B. T.; Althoff, E. A.; Cornish, V. W. *Biochemistry* **2004**, *43*, 10353.
(46) Iannolo, G.; Minenkova, O.; Petruzzelli, R.; Cesareni, G. *J Mol Biol* **1995**, *248*, 835.
(47) Kehoe, J. W.; Kay, B. K. *Chem. Rev.* **2005**, *105*, 4056.
(48) Georgieva, Y.; Konthur, Z. *Molecules* **2011**, *16*, 1667.
(49) Brunet, E.; Chauvin, C.; Choumet, V. r.; Jestin, J.-L. *Nucleic Acids Res* **2002**, *30*, e40.
(50) Crameri, R.; Suter, M. *Gene* **1993**, *137*, 69.
(51) Jiang, J.; Abu-Shilbayeh, L.; Rao, V. B. *Infect Immun* **1997**, *65*, 4770.
(52) Danner, S.; Belasco, J. G. *Proc Natl Acad Sci U S A* **2001**, *98*, 12954.
(53) Castagnoli, L.; Zuccconi, A.; Quondam, M.; Rossi, M.; Vaccaro, P.; Panni, S.; Paoluzi, S.; Santonico, E.; Dente, L.; Cesareni, G. *Comb Chem High Throughput Screening* **2001**, *4*, 121.
(54) Sternberg, N.; Hoess, R. H. *Proc Natl Acad Sci U S A* **1995**, *92*, 1609.
(55) Rosenberg, A. H.; Griffin, K.; Washington, M. T.; Patel, S. S.; Studier, F. W. *J Biol Chem* **1996**, *271*, 26819.
(56) Krumpe, L. R. H.; Atkinson, A. J.; Smythers, G. W.; Kandel, A.; Schumacher, K. M.; McMahon, J. B.; Makowski, L.; Mori, T. *Proteomics* **2006**, *6*, 4210.
(57) Kemp, P.; Garcia, L. R.; Molineux, I. J. *Virology* **2005**, *340*, 307.
(58) Molineux, I. J. *Mol Microbiol* **2001**, *40*, 1.
(59) Molineux, I. J. In *The Bacteriophages*; Calendar, R., Ed.; Oxford University Press: New York City, 2006; Vol. 1, p 277.
(60) Struthers-Schlinke, J. S.; Robins, W. P.; Kemp, P.; Molineux, I. J. *J Mol Biol* **2000**, *301*, 35.
(61) Studier, F. W. *Virology* **1969**, *39*, 562.
(62) Studier, F. W. *Science* **1972**, *176*, 367.
(63) Sche, P. P.; McKenzie, K. M.; White, J. D.; Austin, D. J. *Chem Biol* **1999**, *6*, 707.
(64) Wilson, K. P.; Yamashita, M. M.; Sintchak, M. D.; Rotstein, S. H.; Murcko, M. A.; Boger, J.; Thomson, J. A.; Fitzgibbon, M. J.; Black, J. R.; Navia, M. A. *Acta Crystallogr, Sect D: Biol Crystallogr* **1995**, *51*, 511.
(65) Sche, P. P.; McKenzie, K. M.; White, J. D.; Austin, D. J. *Chem Biol* **2001**, *8*, 399.
(66) Schreuder, M. P.; Brekelmans, S.; Van Den Ende, H.; Klis, F. M. *Yeast* **1993**, *9*, 399.
(67) Kondo, A.; Ueda, M. *Appl Microbiol Biotechnol* **2004**, *64*, 28.
(68) Lipke, P. N.; Kurjan, J. *Microbiol Rev* **1992**, *56*, 180.
(69) Lahti, J. L.; Cochran, J. R. In *Therapeutic Monoclonal Antibodies*; John Wiley & Sons, Inc.: 2009, p 213.
(70) Wadle, A.; Mischo, A.; Imig, J.; Wüllner, B.; Hensel, D.; Wätzig, K.; Neumann, F.; Kubuschok, B.; Schmidt, W.; Old, L. J.; Pfreundschuh, M.; Renner, C. *Int J Cancer* **2005**, *117*, 104.
(71) Bidlingmaier, S.; Liu, B. *Mol Cell Proteomics* **2006**, *5*, 533.
(72) Bidlingmaier, S.; Liu, B. *Methods Mol Biol* **2011**, *729*, 199.
(73) Bidlingmaier, S.; Liu, B. *Methods Mol Biol* **2011**, *729*, 211.
(74) Bidlingmaier, S.; Wang, Y.; Liu, Y.; Zhang, N.; Liu, B. *Mol Cell Proteomics* **2011**, *10*.
(75) Hinds, P. W.; Blake, R. D. *J Biomol Struct Dyn* **1985**, *3*, 543.
(76) Rothe, A.; Hosse, R. J.; Power, B. E. *FASEB J* **2006**, *20*, 1599.
(77) Bradbury, A. R. M.; Sidhu, S.; Dubel, S.; McCafferty, J. *Nat Biotech* **2011**, *29*, 245.
(78) McKenzie, K. M.; Videlock, E. J.; Splittgerber, U.; Austin, D. J. *Angew Chem, Int Ed Engl* **2004**, *43*, 4052.
(79) Jin, Y.; Yu, J.; Yu, Y. G. *Chem Biol* **2002**, *9*, 157.
(80) Shim, J. S.; Lee, J.; Park, H.-J.; Park, S.-J.; Kwon, H. J. *Chem Biol* **2004**, *11*, 1455.
(81) Yamazaki, T.; Aoki, S.; Ohta, K.; Hyuma, S.; Sakaguchi, K.; Sugawara, F. *Bioorg Med Chem Lett* **2004**, *14*, 4343.
(82) Videlock, E. J.; Chung, V. K.; Mohan, M. A.; Strok, T. M.; Austin, D. J. *J Am Chem Soc* **2004**, *126*, 3730.
(83) Videlock, E. J.; Chung, V. K.; Hall, J. M.; Hines, J.; Agapakis, C. M.; Austin, D. J. *J Am Chem Soc* **2005**, *127*, 1600.
(84) Piggott, A. M.; Karuso, P. *Marine Drugs* **2005**, *3*, 36.
(85) Takakusagi, Y.; Ohta, K.; Kuramochi, K.; Morohashi, K.; Kobayashi, S.; Sakaguchi, K.; Sugawara, F. *Bioorg Med Chem Lett* **2005**, *15*, 4846.
(86) Morimura, T.; Noda, N.; Kato, Y.; Watanabe, T.; Saitoh, T.; Yamazaki, T.; Takada, K.; Aoki, S.; Ohta, K.; Ohshige, M. *J Antibiot* **2006**, *59*, 625.

(87) Kim, H.; Deng, L.; Xiong, X.; Hunter, W. D.; Long, M. C.; Pirrung, M. C. *J Med Chem* **2007**, *50*, 3423.
(88) Ladwa, S. R.; Dilly, S. J.; Clark, A. J.; Marsh, A.; Taylor, P. C. *ChemMedChem* **2008**, *3*, 742.
(89) Piggott, A. M.; Karuso, P. *ChemBioChem* **2008**, *9*, 524.
(90) Saitoh, T.; Kuramochi, K.; Imai, T.; Takata, K.; Takehara, M.; Kobayashi, S.; Sakaguchi, K.; Sugawara, F. *Bioorg Med Chem* **2008**, *16*, 5815.
(91) Piggott, A. M.; Kriegel, A. M.; Willows, R. D.; Karuso, P. *Bioorg Med Chem* **2009**, *17*, 6841.
(92) Van Dorst, B.; Mehta, J.; Rouah-Martin, E.; Coen, W. D.; Blust, R.; Robbens, J. *Toxicol in Vitro* **2011**, *25*, 388.
(93) Dilly, S. J.; Bell, M. J.; Clark, A. J.; Marsh, A.; Napier, R. M.; Sergeant, M. J.; Thompson, A. J.; Taylor, P. C. *Chem Commun (Cambridge, U K)* **2007**, 2808.
(94) Savinov, S. N.; Austin, D. J. *Comb Chem High Throughput Screening* **2001**, *4*, 593.
(95) Boehmerle, W.; Splittgerber, U.; Lazarus, M. B.; McKenzie, K. M.; Johnston, D. G.; Austin, D. J.; Ehrlich, B. E. *Proc Natl Acad Sci U S A* **2006**, *103*, 18356.
(96) He, Q. L.; Jiang, H.; Zhang, F.; Chen, H. B.; Tang, G. L. *Bioorg Med Chem Lett* **2008**, *18*, 3995.
(97) Cheng, K.-W.; Wong, C.-C.; Wang, M.; He, Q.-Y.; Chen, F. *Mass Spectrom Rev* **2010**, *29*, 126.
(98) Johnson, K. M.; Chen, X.; Boitano, A.; Swenson, L.; Opipari, A. W. *Chem Biol* **2005**, *12*, 485.
(99) Kuramochi, K.; Miyano, Y.; Enomoto, Y.; Takeuchi, R.; Ishi, K.; Takakusagi, Y.; Saitoh, T.; Fukudome, K.; Manita, D.; Takeda, Y. *Bioconjugate* **2008**, *19*, 2417.
(100) Liu, S.; Sun, Y.; Chen, H.; Song, S.; Xu, Y. *Appl Biochem Biotechnol* **2010**, *162*, 1206.
(101) Takakusagi, Y.; Takakusagi, K.; Sugawara, F.; Sakaguchi, K. *Expert Opinion on Drug Discovery* **2010**, *5*, 361.
(102) Hengerer, A.; Decker, J.; Prohaska, E.; Hauck, S.; Koesslinger, C.; Wolf, H. *Biosensors and Bioelectronics* **1999**, *14*, 139.
(103) Takakusagi, Y.; Takakusagi, K.; Kuramochi, K.; Kobayashi, S.; Sugawara, F.; Sakaguchi, K. *Bioorg Med Chem* **2007**, *15*, 7590.
(104) Morohashi, K.; Yoshino, A.; Yoshimori, A.; Saito, S.; Tanuma, S.; Sakaguchi, K.; Sugawara, F. *Biochem Pharmacol* **2005**, *70*, 37.
(105) Morohashi, K.; Arai, T.; Saito, S.; Watanabe, M.; Sakaguchi, K.; Sugawara, F. *Comb Chem High Throughput Screening* **2006**, *9*, 55.
(106) Gearhart, D. A.; Toole, P. F.; Warren Beach, J. *J Neurosci Res* **2002**, *44*, 255.
(107) Di Marco, A.; Gaetani, M.; Scarpinato, B. *Cancer Chemoth Rep* **1969**, *53*, 33.
(108) Pirrung, M. C.; Liu, Y.; Deng, L.; Halstead, D. K.; Li, Z.; May, J. F.; Wedel, M.; Austin, D. A.; Webster, N. J. G. *J Am Chem Soc* **2005**, *127*, 4609.
(109) Webster, N. J. G.; Park, K.; Pirrung, M. C. *ChemBioChem* **2003**, *4*, 379.
(110) Hamann, M. T.; Scheuer, P. J. *J Am Chem Soc* **1993**, *115*, 5825.
(111) Manita, D.; Toba, Y.; Takakusagi, Y.; Matsumoto, Y.; Kusayanagi, T.; Takakusagi, K.; Tsukuda, S.; Takada, K.; Kanai, Y.; Kamisuki, S.; Sakaguchi, K.; Sugawara, F. *Bioorg Med Chem* **2011**, *19*, 7690.
(112) Takakusagi, Y.; Kuroiwa, Y.; Sugawara, F.; Sakaguchi, K. *Bioorg Med Chem* **2008**, *16*, 7410.
(113) Nishiyama, K.; Takakusagi, Y.; Kusayanagi, T.; Matsumoto, Y.; Habu, S.; Kuramochi, K.; Sugawara, F.; Sakaguchi, K.; Takahashi, H.; Natsugari, H.; Kobayashi, S. *Bioorg Med Chem* **2009**, *17*, 195.
(114) Chen, Y. H.; Du, G. H.; Zhang, J. T. *Acta Pharmacol Sin* **2000**, *21*, 463.
(115) Jung, H. J.; Kim, K. H.; Kim, N. D.; Han, G.; Kwon, H. J. *Bioorg Med Chem Lett* **2011**, *21*, 1052.
(116) Verpoorte, R. *Drug Discovery Today* **1998**, *3*, 232.
(117) Wink, M. *Phytochemistry* **2003**, *64*, 3.
(118) Koehn, F. E.; Carter, G. T. *Nat Rev Drug Discovery* **2005**, *4*, 206.
(119) Harvey, A. *Drug Discovery Today* **2000**, *5*, 294.
(120) Leal, M. C.; Puga, J.; Serodio, J.; Gomes, N. C. M.; Calado, R. *PLoS One* **2012**, *7*, e30580.
(121) Bergmann, W.; Feeney, R. J. *J Org Chem* **1951**, *16*, 981.
(122) Capon, R. J. *Aust J Chem* **2010**, *63*, 851.
(123) Prommer, E. *Drugs Today (Barc)* **2006**, *42*, 369.
(124) Molinski, T. F.; Dalisay, D. S.; Lievens, S. L.; Saludes, J. P. *Nat Rev Drug Discovery* **2009**, *8*, 69.
(125) Mayer, A. M. S.; Glaser, K. B.; Cuevas, C.; Jacobs, R. S.; Kem, W.; Little, R. D.; McIntosh, J. M.; Newman, D. J.; Potts, B. C.; Shuster, D. E. *Trends Pharmacol Sci* **2010**, *31*, 255.
(126) Mayer, A. M. S.; Mayer, A. M. S., Ed. Downers Grove, IL, 2012; Vol. 2012.
(127) US National Institute of Health http://www.clinicaltrials.gov 2012.
(128) Cragg, G. M.; Grothaus, P. G.; Newman, D. J. *Chem. Rev.* **2009**, *109*, 3012.
(129) Cragg, G. M.; Newman, D. J. *Phytochem Rev* **2009**, *8*, 313.

(130) Newman, D. J.; Cragg, G. M. *Spec Publ - R Soc Chem* **2011**, *320*, 3.
(131) Harper, M. K.; Bugni, T. S.; Copp, B. R.; James, R. D.; Lindsay, B. S.; Richardson, A. D.; Schnabel, P. C.; Tasdemir, D.; VanWagoner, R. M.; Verbitski, S. M.; Ireland, C. M. In *Marine Chemical Ecology*; McClintock, J. B., Baker, B. J., Eds.; CRC Press LLC: 2001, p 3.
(132) Newman, D. J.; Cragg, G. M. *Curr Med Chem* **2004**, *11*, 1693.
(133) Newman, D. J.; Cragg, G. M.; Battershill, C. N. *Diving Hyperb Med.* **2009**, *39*, 215.
(134) Capon, R. J. In *Bioactive Compounds from Natural Sources*; 2nd ed.; Tringali, C., Ed.; CRC Press: 2012, p 579.
(135) Love, G. D.; Grosjean, E.; Stalvies, C.; Fike, D. A.; Grotzinger, J. P.; Bradley, A. S.; Kelly, A. E.; Bhatia, M.; Meredith, W.; Snape, C. E.; Bowring, S. A.; Condon, D. J.; Summons, R. E. *Nature* **2009**, *457*, 718.
(136) Srivastava, M.; Simakov, O.; Chapman, J.; Fahey, B.; Gauthier, M. E. A.; Mitros, T.; Richards, G. S.; Conaco, C.; Dacre, M.; Hellsten, U.; Larroux, C.; Putnam, N. H.; Stanke, M.; Adamska, M.; Darling, A.; Degnan, S. M.; Oakley, T. H.; Plachetzki, D. C.; Zhai, Y.; Adamski, M.; Calcino, A.; Cummins, S. F.; Goodstein, D. M.; Harris, C.; Jackson, D. J.; Leys, S. P.; Shu, S.; Woodcroft, B. J.; Vervoort, M.; Kosik, K. S.; Manning, G.; Degnan, B. M.; Rokhsar, D. S. *Nature* **2010**, *466*, 720.
(137) Müller, W. E. G. *Comp Biochem Physiol, Part A: Mol Integr Physiol* **2001**, *129*, 433.
(138) Singh, I. P.; Bodiwala, H. S. *Nat Prod Rep* **2010**, *27*, 1781.
(139) Okunade, A. L.; Elvin-Lewis, M. P. F.; Lewis, W. H. *Phytochemistry* **2004**, *65*, 1017.
(140) Copp, B. R. *Nat Prod Rep* **2003**, *20*, 535.
(141) Copp, B. R.; Pearce, A. N. *Nat Prod Rep* **2007**, *24*, 278.
(142) Garcia, A.; Bocanegra-Garcia, V.; Palma-Nicolas, J. P.; Rivera, G. *Eur J Med Chem* **2012**, *49*, 1.
(143) Hooper, J. N. A. *"Sponguide". Guide to Sponge Collection and Identification*; Queensland Museum: Brisbane, 2000.
(144) Gershenzon, J.; Dudareva, N. *Nat Chem Biol* **2007**, *3*, 408.
(145) Becerro, M. A.; Turon, X.; Uriz, M. J. *J Chem Ecol* **1997**, *23*, 1527.
(146) Wulff, J. L. *Can J Zoolog* **2006**, *84*, 146.
(147) Capon, R. J.; Elsbury, K.; Butler, M. S.; Lu, C. C.; Hooper, J. N. A.; Rostas, J. A. P.; O'Brien, K. J.; Mudge, L. M.; Sim, A. T. R. *Experientia* **1993**, *49*, 263.
(148) Faulkner, D. J. *Nat Prod Rep* **2002**, *19*, 1.
(149) Blunt, J. W.; Copp, B. R.; Keyzers, R. A.; Munro, M. H. G.; Prinsep, M. R. *Nat Prod Rep* **2012**, *29*, 144.
(150) Ebel, R.; Brenzinger, M.; Kunze, A.; Gross, H. J.; Proksch, P. *J Chem Ecol* **1997**, *23*, 1451.
(151) Laport, M. S.; Santos, O. C. S.; Muricy, G. *Curr Pharm Biotechnol* **2009**, *10*, 86.
(152) Rice, L. B. *Am J Infect Control* **2006**, *34*, S11.
(153) Llewellyn, L. E.; Burnell, J. N. *Pesticide Outlook* **2000**, *11*, 64.
(154) Motti, C.; Bourguet-Kondracki, M.-L.; Longeon, A.; Doyle, J.; Llewellyn, L.; Tapiolas, D.; Yin, P. *Molecules* **2007**, *12*, 1376.
(155) Walsh, M. J.; Harrington, R. B.; Powles, S. B. *Crop Science* **2012**, *52*, 1343.
(156) Duke, S. O. *Environ Health Perspect* **1990**, *87*, 263.
(157) Paul, V. J.; Ritson-Williams, R. *Nat Prod Rep* **2008**, *25*, 662.
(158) Paul, V. J.; Ritson-Williams, R.; Sharp, K. *Nat Prod Rep* **2011**, *28*, 345.
(159) Faulkner, D. J. *Nat Prod Rep* **2000**, *17*, 1.
(160) Fontana, A.; Avila, C.; Martinez, E.; Ortea, J.; Trivellone, E.; Cimino, G. *J Chem Ecol* **1993**, *19*, 339.
(161) Becerro, M. A.; Goetz, G.; Paul, V. J.; Scheuer, P. J. *J Chem Ecol* **2001**, *27*, 2287.
(162) Hill, R. T. B., MD, US), Enticknap, Julie (Baltimore, MD, US), Rao, Karumanchi V. (Oxford, MS, US), Hamann, Mark T. (Oxford, MS, US) In *US Patent Application Publications*; Publications, U. P. A., Ed. United States, 2007.
(163) Schmitt, S.; Tsai, P.; Bell, J.; Fromont, J.; Ilan, M.; Lindquist, N.; Perez, T.; Rodrigo, A.; Schupp, P. J.; Vacelet, J.; Webster, N.; Hentschel, U.; Taylor, M. W. *ISME Journal* **2012**, *6*, 564.
(164) Taylor, M. W.; Schupp, P. J.; Dahllof, I.; Kjelleberg, S.; Steinberg, P. D. *Environ Microbiol* **2004**, *6*, 121.
(165) Taylor, M. W.; Radax, R.; Steger, D.; Wagner, M. *Microbiol Mol Biol Rev* **2007**, *71*, 295.
(166) Bergquist, P. R.; Cook, S. d. C. In *Systema Porifera: A Guide to the Classification of Sponges*; Hooper, J., Van Soest, R., Willenz, P., Eds.; Springer Us: 2002, p 1081.
(167) Cimino, G.; De Rosa, S.; De Stefano, S.; Spinella, A.; Sodano, G. *Tetrahedron Lett* **1984**, *25*, 2925.
(168) Fattorusso, E.; Forenza, S.; Minale, L.; Sodano, G. *Gazz Chim Ital* **1971**, *101*, 104.
(169) Thoms, C.; Wolff, M.; Padmakumar, K.; Ebel, R.; Proksch, P. *Z Naturforsch, C: Biosci* **2004**, *59*, 113.

(170) Bergquist, P. R.; Wells, R. J. In *Marine Natural Products*; Scheuer, P. J., Ed. New York, 1983; Vol. 1.
(171) Diers, J. A.; Pennaka, H. K.; Peng, J.; Bowling, J. J.; Duke, S. O.; Hamann, M. T. *J Nat Prod* **2004**, *67*, 2117.
(172) Tsukamoto, S.; Kato, H.; Hirota, H.; Fusetani, N. *Tetrahedron* **1996**, *52*, 8186.
(173) Hamann, M. T.; Scheuer, P. J.; Kelly-Borges, M. *J Org Chem* **1993**, *58*, 6565.
(174) Peng, J.; Li, J.; Hamann, M. T. In *The Alkaloids: Chemistry and Biology*; Cordell, G. A., Ed.; Academic Press: 2005; Vol. Volume 61, p 59.
(175) Blunt, J. W.; Copp, B. R.; Munro, M. H. G.; Northcote, P. T.; Prinsep, M. R. *Nat Prod Rep* **2011**, *28*, 196.
(176) Ebada, S. S.; Edrada, R. A.; Lin, W.; Proksch, P. *Nat Protoc* **2008**, *3*, 1820.
(177) Bauer, A. W.; Kirby, W. M. M.; Sherris, J. C.; Turck, M. *Am J Clin Path* **1966**, *45*, 425.
(178) Mahasneh, A. M. *Phytother Res* **2002**, *16*, 751.
(179) Schantz, E. J.; Lauffer, M. A. *Biochemistry* **1962**, *1*, 658.
(180) Eloff, J. *Planta Med* **1998**, *64*, 711.
(181) Mosmann, T. *J Immunol Methods* **1983**, *65*, 63.
(182) Stevens, M. G.; Kehrli Jr, M. E.; Canning, P. C. *Vet Immunol Immunopath* **1991**, *28*, 45.
(183) Grare, M.; Fontanay, S. p.; Cornil, C.; Finance, C.; Duval, R. l. E. *J Microbiol Methods* **2008**, *75*, 156.
(184) Appendino, G.; Gibbons, S.; Giana, A.; Pagani, A.; Grassi, G.; Stavri, M.; Smith, E.; Rahman, M. *J Nat Prod* **2008**, *71*, 1427.
(185) Cherigo, L.; Pereda-Miranda, R.; Gibbons, S. *Phytochemistry* **2009**, *70*, 222.
(186) Pfaller, M. A.; Messer, S. A.; Coffmann, S. *J Clin Microbiol* **1995**, *33*, 1094.
(187) Pfaller, M. A.; Sheehan, D. J.; Rex, J. H. *Clin Microbiol Rev* **2004**, *17*, 268.
(188) Matar, M. J.; Ostrosky-Zeichner, L.; Paetznick, V. L.; Rodriguez, J. R.; Chen, E.; Rex, J. H. *Antimicrob Agents Chemother* **2003**, *47*, 1647.
(189) United States Environmental Protection Agency In *Aquatic Plant Toxicity Test Using Lemna spp., Tiers I and II* New York, 1996, p 10.
(190) Jones, R. N. *Clinical Infectious Diseases* **2010**, *51*, S81.
(191) Kuete, V.; Ngameni, B.; Simo, C. C. F.; Tankeu, R. K.; Ngadjui, B. T.; Meyer, J. J. M.; Lall, N.; Kuiate, J. R. *J Ethnopharmacol* **2008**, *120*, 17.
(192) Andrews, J. M.; Howe, R. A. *J Antimicrob Chemother* **2011**, *66*, 2726.
(193) Gardner, S. C.; Grue, C. E. *Environ Toxicol Chem* **1996**, *15*, 441.
(194) Lockhart, L. W.; Billeck, B. N.; Baron, C. L. *Hydrobiologia* **1989**, *188*, 353.
(195) Van Soest, R. W. M.; Boury-Esnault, N.; Hooper, J. N. A.; Rützler, K.; de Voogd, N. J.; Alvarez de Glasby, B.; Hajdu, E.; Pisera, A. B.; Manconi, R.; Schoenberg, C.; Janussen, D.; Tabachnick, K. R.; Klautau, M.; Picton, B.; Kelly, M.; Vacelet, J. 2012; Vol. 2012.
(196) Capon, Robert J. *Eur J Org Chem* **2001**, *2001*, 633.
(197) Luna-Herrera, J.; Costa, M. C.; Gonzalez, H. G.; Rodrigues, A. I.; Castilho, P. C. *J Antimicrob Chemother* **2007**, *59*, 548.
(198) Deharo, E.; Ginsburg, H. *Malaria Journal* **2011**, *10*, S5.
(199) Kubo, I.; Muroi, H.; Himejima, M. *J Nat Prod* **1992**, *55*, 1436.
(200) Benharref, A.; Pais, M.; Debitus, C. *J Nat Prod* **1996**, *59*, 177.
(201) Cimino, G.; Rosa, S. D.; Stefano, S. D.; Self, R.; Sodano, G. *Tetrahedron Lett* **1983**, *24*, 3029.
(202) Tsuda, M.; Sakuma, Y.; Kobayashi, J. i. *J Nat Prod* **2001**, *64*, 980.
(203) Morris, S. A.; Andersen, R. J. *Can J Chem* **1989**, *67*, 677.
(204) Xynas, R.; Capon, R. J. *Aust J Chem* **1989**, *42*, 1427.
(205) Pakrashi, S. C.; Achari, B.; Dutta, P. K.; Chakrabarti, A. K.; Giri, C.; Saha, S.; Basa, S. C. *Tetrahedron* **1994**, *50*, 12009.
(206) Nakamura, H.; Wu, H.; Kobayashi, J.; Nakamura, Y.; Ohizumi, Y.; Hirata, Y. *Tetrahedron Lett* **1985**, *26*, 4517.
(207) Kijjoa, A.; Bessa, J.; Wattanadilok, R.; Sawangwong, P.; Nascimento, M. S. J.; Pedro, M.; Silva, A. M. S.; Eaton, G.; van, S. R.; Herz, W. *Z Naturforsch, B: Chem Sci* **2005**, *60*, 904.
(208) Jang, J.-H.; Van Soest, R. W. M.; Fusetani, N.; Matsunaga, S. *J Org Chem* **2007**, *72*, 1217.
(209) Buchanan, M. S.; Carroll, A. R.; Fechner, G. A.; Boyle, A.; Simpson, M.; Addepalli, R.; Avery, V. M.; Hooper, J. N. A.; Cheung, T.; Chen, H.; Quinn, R. J. *J Nat Prod* **2008**, *71*, 1066.
(210) Kalaitzis, J. A.; De, A. L. P.; Hooper, J. N. A.; Quinn, R. J. *Nat Prod Res* **2008**, *22*, 1257.
(211) Ma, K.; Yang, Y.; Deng, Z.; de Voogd, N. J.; Proksch, P.; Lin, W. *Chem Biodiversity* **2008**, *5*, 1313.
(212) Teruya, T.; Iwasaki, A.; Suenaga, K. *Bull Chem Soc* **2008**, *81*, 1026.

(213) Fujiwara, T.; Hwang, J.-H.; Kanamoto, A.; Nagai, H.; Takagi, M.; Shin-ya, K. *J Antibiot* **2009**, *62*, 393.
(214) McCulloch, M. W. B.; Coombs, G. S.; Banerjee, N.; Bugni, T. S.; Cannon, K. M.; Harper, M. K.; Veltri, C. A.; Virshup, D. M.; Ireland, C. M. *Bioorg Med Chem* **2009**, *17*, 2198.
(215) Mukai, H.; Kubota, T.; Aoyama, K.; Mikami, Y.; Fromont, J.; Kobayashi, J. i. *Bioorg Med Chem Lett* **2009**, *19*, 1337.
(216) Feng, Y.; Davis, R. A.; Sykes, M. L.; Avery, V. M.; Camp, D.; Quinn, R. J. *Tetrahedron Lett* **2010**, *51*, 4847.
(217) Kon, Y.; Kubota, T.; Shibazaki, A.; Gonoi, T.; Kobayashi, J. i. *Bioorg Med Chem Lett* **2010**, *20*, 4569.
(218) Xu, M.; Andrews, K. T.; Birrell, G. W.; Tran, T. L.; Camp, D.; Davis, R. A.; Quinn, R. J. *Bioorg Med Chem Lett* **2011**, *21*, 846.
(219) Yin, S.; Davis, R. A.; Shelper, T.; Sykes, M.; Avery, V. M.; Elofsson, M.; Sundin, C.; Quinn, R. *Org Biomol Chem* **2011**.
(220) Bax, A.; Summers, M. F. *J Am Chem Soc* **1986**, *108*, 2093.
(221) Jurek, J.; Yoshida, W. Y.; Scheuer, P. J.; Kelly-Borges, M. *J Nat Prod* **1993**, *56*, 1609.
(222) Arabshahi, L.; Schmitz, F. J. *J Org Chem* **1987**, *52*, 3584.
(223) Fusetani, N.; Masuda, Y.; Nakao, Y.; Matsunaga, S.; van Soest, R. W. M. *Tetrahedron* **2001**, *57*, 7507.
(224) Kobayashi, J. i.; Honma, K.; Sasaki, T.; Tsuda, M. *Chem Pharm Bull* **1995**, *43*, 403.
(225) Tilvi, S.; Rodrigues, C.; Naik, C. G.; Parameswaran, P. S.; Wahidhulla, S. *Tetrahedron* **2004**, *60*, 10207.
(226) Clinical and Laboratory Standards Institute In *M07-A8 Vol. 29 No. 2*; Clinical and Laboratory Standards Institute: 2009.
(227) Bugni, T.; Harper, M.; McCulloch, M.; Reppart, J.; Ireland, C. *Molecules* **2008**, *13*, 1372.
(228) Motti, C. A.; Freckelton, M. L.; Tapiolas, D. M.; Willis, R. H. *J Nat Prod* **2009**, *72*, 290.
(229) Colson, K. L.; Motti, C. A.; Neidig, P.; Fischer, C.; Hicks, J. M.; Tapiolas, D. *Planta Med* **2008**, *74*, SL3.
(230) Simmons, T. L.; Andrianasolo, E.; McPhail, K.; Flatt, P.; Gerwick, W. H. *Mol Cancer Ther* **2005**, *4*, 333.
(231) Thoms, C.; Ebel, R.; Proksch, P. *Prog Mol Subcell Biol* **2006**, *43*, 261.
(232) Willan, R. C. *Journal of the Malacological Society of Australia* **1984**, *6*, 125.
(233) Thoms, C.; Ebel, R.; Hentschel, U.; Proksch, P. *Z Naturforsch, C: Biosci* **2003**, *58*, 426.
(234) Thoms, C.; Schupp, P. J. *J Chem Ecol* **2008**, *34*, 1242.
(235) Becerro, M. A.; Turon, X.; Uriz, M. J.; Templado, J. *Biol J Linn Soc* **2003**, *78*, 429.
(236) Teeyapant, R.; Kreis, P.; Wray, V.; Witte, L.; Proksch, P. *Z Naturforsch, C: Biosci* **1993**, *48*, 640.
(237) Ebel, R.; Marin, A.; Proksch, P. *Biochem Syst Ecol* **1999**, *27*, 769.
(238) Proksch, P.; Putz, A.; Ortlepp, S.; Kjer, J.; Bayer, M. *Phytochem Rev* **2010**, *9*, 475.
(239) Merrifield, R. B. *J Am Chem Soc* **1963**, *85*, 2149.
(240) Guillier, F.; Orain, D.; Bradley, M. *Chem. Rev.* **2000**, *100*, 2091.
(241) Wells, N. J.; Davies, M.; Bradley, M. *J Org Chem* **1998**, *63*, 6430.
(242) Sugimoto, H.; Oda, S.-i.; Otsuki, T.; Hino, T.; Yoshida, T.; Shiro, Y. *Proc Natl Acad Sci U S A* **2006**, *103*, 2611.
(243) Smith, J. R., The Active Site in Human Recombinant Indoleamine-2,3-Dioxygenase (Pers. Comm.).
(244) Smith, J. R.; Evans, K. J.; Wright, A.; Willows, R. D.; Jamie, J. F.; Griffith, R. *Bioorg Med Chem* **2012**, *20*, 1354.
(245) Onodera, K.; Hirano, S.; Kashimura, N. *J Am Chem Soc* **1965**, *87*, 4651.
(246) Bowden, K.; Heilbron, I. M.; Jones, E. R. H.; Weedon, B. C. L. *Journal of the Chemical Society (Resumed)* **1946**, 39.
(247) Hendrickson, W. A.; Pahler, A.; Smith, J. L.; Satow, Y.; Merritt, E. A.; Phizackerley, R. P. *Proc Natl Acad Sci U S A* **1989**, *86*, 2190.
(248) Stanley, M.; Martin, S. R.; Birge, M.; Carbain, B.; Streicher, H. *Org Biomol Chem* **2011**, *9*.
(249) Meldal, M.; Tornøe, C. W. *Chem. Rev.* **2008**, *108*, 2952.
(250) Menchikov, L. G.; Vorogushin, A. V.; Korneva, O. S.; Nefedov, O. M. *Mendeleev Commun* **1995**, *5*, 223.
(251) Tanaka, K.; Kageyama, C.; Fukase, K. *Tetrahedron Lett* **2007**, *48*, 6475.
(252) Baldwin, J. E.; Whitehead, R. C. *Tetrahedron Lett* **1992**, *33*, 2059.
(253) Sakai, R.; Higa, T.; Jefford, C. W.; Bernardinelli, G. *J Am Chem Soc* **1986**, *108*, 6404.

(254) Ang, K.; Holmes, M.; Higa, T.; Hamann, M.; Kara, U. *Antimicrob Agents Chemother* **2000**, *44*, 1645.
(255) Urban, S.; Hickford, S.; Blunt, J.; Munro, M. *Curr Org Chem* **2000**, *4*, 765.
(256) Nakamura, H.; Deng, S.; Kobayashi, J.; Ohizumi, Y.; Tomotake, Y.; Matsuzaki, T.; Hirata, Y. *Tetrahedron Lett* **1987**, *28*, 621.
(257) Peng, J.; Shen, X.; El Sayed, K. A.; Dunbar, D. C.; Perry, T. L.; Wilkins, S. P.; Hamann, M. T.; Bobzin, S.; Huesing, J.; Camp, R.; Prinsen, M.; Krupa, D.; Wideman, M. A. *J Agric Food Chem* **2003**, *51*, 2246.
(258) Edrada, R. A.; Proksch, P.; Wray, V.; Witte, L.; Müller, W. E. G.; Van Soest, R. W. M. *J Nat Prod* **1996**, *59*, 1056.
(259) El Sayed, K. A.; Dunbar, D. C.; Perry, T. L.; Wilkins, S. P.; Hamann, M. T.; Greenplate, J. T.; Wideman, M. A. *J Agric Food Chem* **1997**, *45*, 2735.
(260) Mayer, A. M. S.; Gunasekera, M.; Pomponi, S. A.; Sennett, S. In *PCT International Patent Application*; Harbor Branch Oceanographic Institution, Inc., USA; Midwestern University . 2000, p 33.
(261) Peng, J.; Hu, J.-F.; Kazi, A. B.; Li, Z.; Avery, M.; Peraud, O.; Hill, R. T.; Franzblau, S. G.; Zhang, F.; Schinazi, R. F.; Wirtz, S. S.; Tharnish, P.; Kelly, M.; Wahyuono, S.; Hamann, M. T. *J Am Chem Soc* **2003**, *125*, 13382.
(262) Hamann, M.; Rao, K. V.; Peng, J. In *US Patent Application Publications*; Publications, U. P. A., Ed. 2005.
(263) El Sayed, K. A.; Kelly, M.; Kara, U. A. K.; Ang, K. K. H.; Katsuyama, I.; Dunbar, D. C.; Khan, A. A.; Hamann, M. T. *J Am Chem Soc* **2001**, *123*, 1804.
(264) Hamann, M.; El Sayed, K. A. In *PCT International Patent Application*; Application, P. I. P., Ed. 2002; Vol. WO0217917 (A1), p 1.
(265) Hamann, M.; Alonso, D.; Martín-Aparicio, E.; Fuertes, A.; Pérez-Puerto, M.; Castro, A.; Morales, S.; Navarro, M.; del Monte-Millán, M.; Medina, M. *J Nat Prod* **2007**, *70*, 1397.
(266) Cohen, P.; Frame, S. *Nat Rev Mol Cell Biol* **2001**, *2*, 769.
(267) Osolodkin, D.; Shulga, D.; Palyulin, V.; Zefirov, N. *Russ Chem Bull* **2010**, *59*, 1983.
(268) Coghlan, M. P.; Culbert, A. A.; Cross, D. A. E.; Corcoran, S. L.; Yates, J. W.; Pearce, N. J.; Rausch, O. L.; Murphy, G. J.; Carter, P. S.; Roxbee Cox, L.; Mills, D.; Brown, M. J.; Haigh, D.; Ward, R. W.; Smith, D. G.; Murray, K. J.; Reith, A. D.; Holder, J. C. *Chem Biol* **2000**, *7*, 793.
(269) Peng, J.; Rao, K. V.; Choo, Y.-M.; Hamann, M. T. In *Modern Alkaloids*; Wiley-VCH Verlag GmbH & Co. KGaA: 2008, p 189.
(270) Higa, T.; Sakai, R.; Publications, U. P. A., Ed.; Harbor Branch Oceanographic Institution, Inc., USA . 1988, p 27.
(271) Ichiba, T.; Corgiat, J. M.; Scheuer, P. J.; Kelly-Borges, M. *J Nat Prod* **1994**, *57*, 168.
(272) Rao, K. V.; Kasanah, N.; Wahyuono, S.; Tekwani, B. L.; Schinazi, R. F.; Hamann, M. T. *J Nat Prod* **2004**, *67*, 1314.
(273) Winkler, J. D.; Londregan, A. T.; Hamann, M. T. *Org Lett* **2006**, *8*, 2591.
(274) Winkler, J. D.; Londregan, A. T.; Hamann, M. T. *Org Lett* **2007**, *9*, 4467.
(275) Rao, K.; Donia, M.; Peng, J.; Garcia-Palomero, E.; Alonso, D.; Martinez, A.; Medina, M.; Franzblau, S.; Tekwani, B.; Khan, S. *J Nat Prod* **2006**, *69*, 1034.
(276) Boursereau, Y.; Coldham, I. *Bioorg Med Chem Lett* **2004**, *14*, 5841.
(277) Flessner, T.; Doye, S. *J Prakt Chem* **1999**, *341*, 186.
(278) Chan, T. R.; Hilgraf, R.; Sharpless, K. B.; Fokin, V. V. *Org Lett* **2004**, *6*, 2853.
(279) Debono, M.; Barnhart, M.; Carrell, C.; Hoffmann, J.; Occolowitz, J.; Abbott, B.; Fukuda, D.; Hamill, R.; Biemann, K.; Herlihy, W. *J Antibiot* **1987**, *40*, 761.
(280) Sauermann, R.; Rothenburger, M.; Graninger, W.; Joukhadar, C. *Pharmacology* **2008**, *81*, 79.
(281) Vilhena, C.; Bettencourt, A. *Mini Reviews in Medicinal Chemistry* **2012**, *12*, 202.
(282) Boaretti, M.; Canepari, P. *Antimicrob Agents Chemother* **1995**, *39*, 2068.
(283) Boaretti, M.; Canepari, P.; del Mar Lleo, M.; Satta, G. *J Antimicrob Chemother* **1993**, *31*, 227.
(284) Canepari, P.; Boaretti, M.; del Mar Lleo, M.; Satta, G. *Antimicrob Agents Chemother* **1990**, *34*, 1220.
(285) Laganas, V.; Alder, J.; Silverman, J. A. *Antimicrob Agents Chemother* **2003**, *47*, 2682.
(286) Silverman, J. A.; Perlmutter, N. G.; Shapiro, H. M. *Antimicrob Agents Chemother* **2003**, *47*, 2538.
(287) Jung, D.; Rozek, A.; Okon, M.; Hancock, R. E. W. *Chem Biol* **2004**, *11*, 949.
(288) Scott, W. R. P.; Baek, S.-B.; Jung, D.; Hancock, R. E. W.; Straus, S. K. *BBA-Biomembranes* **2007**, *1768*, 3116.
(289) Robbel, L.; Marahiel, M. A. *J Biol Chem* **2010**, *285*, 27501.

(290) Jones, T.; Yeaman, M. R.; Sakoulas, G.; Yang, S.-J.; Proctor, R. A.; Sahl, H.-G.; Schrenzel, J.; Xiong, Y. Q.; Bayer, A. S. *Antimicrob Agents Chemother* **2008**, *52*, 269.
(291) Hobbs, J. K.; Miller, K.; O'Neill, A. J.; Chopra, I. *J Antimicrob Chemother* **2008**, *62*, 1003.
(292) Muthaiyan, A.; Silverman, J. A.; Jayaswal, R. K.; Wilkinson, B. J. *Antimicrob Agents Chemother* **2008**, *52*, 980.
(293) Muraih, J. K.; Pearson, A.; Silverman, J.; Palmer, M. *Biochimica et Biophysica Acta - Biomembranes* **2011**, *1808*, 1154.
(294) Fenton, C.; Keating, G. M.; Curran, M. P. *Drugs* **2004**, *64*, 445.
(295) Echevarria, K.; Datta, P.; Cadena, J.; Lewis, J. S. *J Antimicrob Chemother* **2005**, *55*, 599.
(296) Tally, F. P.; Zeckel, M.; Wasilewski, M. M.; Carini, C.; Berman, C. L.; Drusano, G. L.; Oleson JR, F. B. *Expert Opin Invest Drugs* **1999**, *8*, 1223.
(297) Kostrominova, T.; Coleman, S.; Oleson, F.; Faulkner, J.; Larkin, L. *In Vitro Cell Dev-An* **2010**, *46*, 613.
(298) Paolini, G. V.; Shapland, R. H. B.; Van Hoorn, W. P.; Mason, J. S.; Hopkins, A. L. *Nat Biotechnol* **2006**, *24*, 805.
(299) Baltz, R. H.; Miao, V.; Wrigley, S. K. *Nat Prod Rep* **2005**, *22*, 717.
(300) Counter, F.; Allen, N.; Fukuda, D.; Hobbs, J.; Ott, J.; Ensminger, P.; Mynderse, J.; Preston, D.; Wu, C. *J Antibiot* **1990**, *43*, 616.
(301) Ball, L.-J.; Goult, C. M.; Donarski, J. A.; Micklefield, J.; Ramesh, V. *Org Biomol Chem* **2004**, *2*, 8.
(302) Rotondi, K. S.; Gierasch, L. M. *Peptide Science* **2005**, *80*, 374.
(303) Ho, S.; Jung, D.; Calhoun, J.; Lear, J.; Okon, M.; Scott, W.; Hancock, R.; Straus, S. *European Biophysics Journal* **2008**, *37*, 421.
(304) Jung, D.; Powers, J. P.; Straus, S. K.; Hancock, R. E. W. *Chem Phys Lipids* **2008**, *154*, 120.
(305) Woodworth, J. R.; Nyhart, E. H.; Brier, G. L.; Wolny, J. D.; Black, H. R. *Antimicrob Agents Chemother* **1992**, *36*, 318.
(306) Nguyen, K. T.; Ritz, D.; Gu, J.-Q.; Alexander, D.; Chu, M.; Miao, V.; Brian, P.; Baltz, R. H. *Proc Natl Acad Sci U S A* **2006**, *103*, 17462.
(307) Siedlecki, J.; Hill, J.; Parr, I.; Yu, X.; Morytko, M.; Zhang, Y.; Silverman, J.; Controneo, N.; Laganas, V.; Li, T. *Bioorg Med Chem Lett* **2003**, *13*, 4245.
(308) National Committee for Clinical Laboratory Standards **2009**, *29*.
(309) Muangsiri, W.; Kearney, W.; Teesch, L.; Kirsch, L. *Int J Pharm* **2005**, *289*, 133.
(310) Pu, Y. J.; Vaid, R. K.; Boini, S. K.; Towsley, R. W.; Doecke, C. W.; Mitchell, D. *Org Process Res Dev* **2008**, *13*, 310.
(311) Liu, W.; Greytak, A. B.; Lee, J.; Wong, C. R.; Park, J.; Marshall, L. F.; Jiang, W.; Curtin, P. N.; Ting, A. Y.; Nocera, D. G.; Fukumura, D.; Jain, R. K.; Bawendi, M. G. *J Am Chem Soc* **2010**, *132*, 472.
(312) Gilles, M. A.; Hudson, A. Q.; Borders, C. L. *Anal Biochem* **1990**, *184*, 244.
(313) Abdel-Magid, A. F.; Mehrman, S. J. *Org Process Res Dev* **2006**, *10*, 971.
(314) Liu, F.-T.; Leonard, N. J. *J Am Chem Soc* **1979**, *101*, 996.
(315) Upadhya, K.; Khattak, I. K.; Mullah, B. *Nucleos Nucleot Nucl* **2005**, *24*, 919.
(316) Zhou, X.; Shearer, J.; Rokita, S. E. *J Am Chem Soc* **2000**, *122*, 9046.
(317) Mock, D. M.; Lankford, G. L.; Cazin, J. *J Nutr* **1993**, *123*, 1844.
(318) Johnson, A. P.; Mushtaq, S.; Warner, M.; Livermore, D. M. *Int J Antimicrob Agents* **2004**, *24*, 315.
(319) Klayman, D. L. *Science* **1985**, *v228*, p1049(7).
(320) Liu, J. M.; Ni, M. Y.; Fan, J. F.; Tu, Y. Y.; Wu, Y. L.; Chou, W. S.; Coordinating Group for Research on the Structure of Qing Hau Sau *Kexue Tongbao* **1977**, *22*, 142.
(321) Efferth, T. *Planta Med* **2007**, *73*, 299.
(322) Efferth, T. *Curr Drug Targets* **2006**, *7*, 407.
(323) Woodrow, C. J.; Bustamante, L. Y. *Trends Parasitol* **2011**, *27*, 2.
(324) Krishna, S.; Pulcini, S.; Fatih, F.; Staines, H. *Trends Parasitol* **2010**, *26*, 517.
(325) Uhlemann, A.-C.; Cameron, A.; Eckstein-Ludwig, U.; Fischbarg, J.; Iserovich, P.; Zuniga, F. A.; East, M.; Lee, A.; Brady, L.; Haynes, R. K.; Krishna, S. *Nat Struct Mol Biol* **2012**, *19*, 264.
(326) Uhlemann, A. C.; Cameron, A.; Eckstein-Ludwig, U.; Fischbarg, J.; Iserovich, P.; Zuniga, F. A.; East, M.; Lee, A.; Brady, L.; Haynes, R. K.; Krishna, S. *Nat Struct Mol Biol* **2005**, *12*, 9.
(327) Krishna, S.; Pulcini, S.; Fatih, F. A.; Staines, H. M. *Trends Parasitol* **2011**, *27*, 3.
(328) Hamacher-Brady, A.; Stein, H. A.; Turschner, S.; Toegel, I.; Mora, R.; Jennewein, N.; Efferth, T.; Eils, R.; Brady, N. R. *J Biol Chem* **2011**, *286*, 6587.
(329) Krishna, S.; Bustamante, L.; Haynes, R. K.; Staines, H. M. *Trends Pharmacol Sci* **2008**, *29*, 520.
(330) World Health Organisation *Guidelines for the Treatment of Malaria - 2nd Edition*; 2nd ed.; World Health Organisation, 2010.
(331) Wang, J.; Huang, L.; Li, J.; Fan, Q.; Long, Y.; Li, Y.; Zhou, B. *PLoS One* **2010**, *5*, e9582.

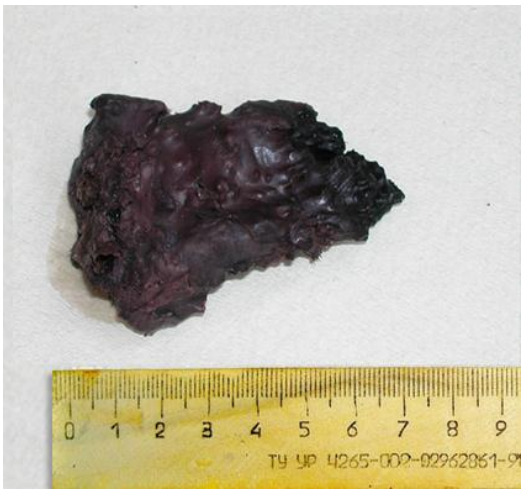
(332) Klonis, N.; Crespo-Ortiz, M. P.; Bottova, I.; Abu-Bakar, N.; Kenny, S.; Rosenthal, P. J.; Tilley, L. *Proc Natl Acad Sci U S A* **2011**, *108*, 11405.
(333) Coghi, P.; Basilico, N.; Taramelli, D.; Chan, W.-C.; Haynes, R. K.; Monti, D. *ChemMedChem* **2009**, *4*, 2045.
(334) Maeno, Y.; Toyoshima, T.; Fujioka, H.; Ito, Y.; Meshnick, S. R.; Benakis, A.; Milhous, W. K.; Aikawa, M. *The American Journal of Tropical Medicine and Hygiene* **1993**, *49*, 485.
(335) Krishna, S.; Woodrow, C. J.; Staines, H. M.; Haynes, R. K.; Mercereau-Puijalon, O. *Trends Mol Med* **2006**, *12*, 200.
(336) Toovey, S.; Bustamante, L. Y.; Uhlemann, A.-C.; East, J. M.; Krishna, S. *Basic Clin Pharmacol Toxicol* **2008**, *103*, 209.
(337) Riganti, C.; Doublier, S.; Viarisio, D.; Miraglia, E.; Pescarmona, G.; Ghigo, D.; Bosia, A. *Br J Pharmacol* **2009**, *156*, 1054.
(338) Nagamune, K.; Beatty, W. L.; Sibley, L. D. *Eukaryotic Cell* **2007**, *6*, 2147.
(339) Mishina, Y. V.; Krishna, S.; Haynes, R. K.; Meade, J. C. *Antimicrob Agents Chemother* **2007**, *51*, 1852.
(340) Moore, C. M.; Hoey, E. M.; Trudgett, A.; Timson, D. J. *FEMS Yeast Res* **2011**, *11*, 233.
(341) Crespo-Ortiz, M. P.; Wei, M. Q. *J Biomed Biotechnol* **2012**, *2012*.
(342) Zhou, H. J.; Wang, W. Q.; Wu, G. D.; Lee, J.; Li, A. *Vasc Pharmacol* **2007**, *47*, 131.
(343) Anfosso, L.; Efferth, T.; Albini, A.; Pfeffer, U. *Pharmacogenomics J* **2006**, *6*, 269.
(344) Longo, M.; Zanoncelli, S.; Manera, D.; Brughera, M.; Colombo, P.; Lansen, J.; Mazue, G.; Gomes, M.; Taylor, W. R. J.; Olliaro, P. *Reprod Toxicol* **2006**, *21*, 83.
(345) Dell'Eva, R.; Pfeffer, U.; Vene, R.; Anfosso, L.; Forlani, A.; Albini, A.; Efferth, T. *Biochem Pharmacol* **2004**, *68*, 2359.
(346) Huan-Huan, C.; Li-Li, Y.; Shang-Bin, L. *Cancer Letters* **2004**, *211*, 163.
(347) Chen, H.-H.; Zhou, H.-J.; Wang, W.-Q.; Wu, G.-D. *Cancer Chemoth Pharm* **2004**, *53*, 423.
(348) Chen, H. H.; Zhou, H. J.; Fang, X. *Pharmacol Res* **2003**, *48*, 231.
(349) Wu, G. D.; Zhou, H. J.; Wu, X. H. *Vasc Pharmacol* **2004**, *41*, 205.
(350) Huang, X.-J.; Ma, Z.-Q.; Zhang, W.-P.; Lu, Y.-B.; Wei, E.-Q. *J Pharm Pharmacol* **2007**, *59*, 849.
(351) Haynes, Richard K.; Chan, H.-W.; Cheung, M.-K.; Lam, W.-L.; Soo, M.-K.; Tsang, H.-W.; Voerste, A.; Williams, Ian D. *Eur J Org Chem* **2002**, *2002*, 113.
(352) Woerdenbarg, H. J.; Moskal, T. A.; Pras, N.; Malingre, T. M. *J Nat Prod* **1993**, *56*, 856.
(353) Lai, H.; Singh, N. P. *Cancer Letters* **1995**, *91*, 41.
(354) Baker, J. K.; McChesney, J. D.; Chi, H. T. *Pharm Res* **1993**, *10*, 662.
(355) Drew, M. G. B.; Metcalfe, J.; Dascombe, M. J.; Ismail, F. M. D. *J Med Chem* **2006**, *49*, 6065.
(356) Batty, K. T.; Ilett, K. F.; Davis, T.; Davis, M. E. *J Pharm Pharmacol* **1996**, *48*, 22.
(357) Luo, X.-D.; Yeh, H. J. C.; Brossi, A.; Flippen-Anderson, J. L.; Gillardi, R. *Helv Chim Acta* **1984**, *67*, 1515.
(358) Cho, S.; Oh, S.; Um, Y.; Jung, J.-H.; Ham, J.; Shin, W.-S.; Lee, S. *Bioorg Med Chem Lett* **2009**, *19*, 382.
(359) Bousejra-El Garah, F.; Pitie, M.; Vendier, L.; Meunier, B.; Robert, A. *J Biol Inorg Chem* **2009**, *14*, 601.
(360) Degani, M.; Narkhede, S.; Pedgaonkar, Y.; Chavan, S.; Sharma, M. K.; Organisation, W. I. P., Ed.; Almet Corporation: International, 2008; Vol. WO 2008/087667 At, p 19.
(361) Siekierka, J. J.; Hung, S. H. Y.; Poe, M.; Lin, C. S.; Sigal, N. H. *Nature* **1989**, *341*, 755.
(362) Harding, M. W.; Galat, A.; Uehling, D. E.; Schreiber, S. L. *Nature* **1989**, *341*, 758.
(363) Kino, T.; Hatanaka, H.; Hashimoto, M.; Nishiyama, M.; Goto, T.; Okuhara, M.; Kosaka, M.; Aoki, H.; Imanaka, H. *J Antibiot* **1987**, *40*, 1249.
(364) Kang, C. B.; Hong, Y.; Dhe-Paganon, S.; Yoon, H. S. *Neurosignals* **2008**, *16*, 318.
(365) Kay, J. E. *Biochem J* **1996**, *314*, 361.
(366) Zeytuni, N.; Zarivach, R. *Structure* **2012**, *20*, 397.
(367) Peng, J.; Kudrimoti, S.; Prasanna, S.; Odde, S.; Doerksen, R. J.; Pennaka, H. K.; Choo, Y.-M.; Rao, K. V.; Tekwani, B. L.; Madgula, V.; Khan, S. I.; Wang, B.; Mayer, A. M. S.; Jacob, M. R.; Tu, L. C.; Gertsch, J.; Hamann, M. T. *J Med Chem* **2010**, *53*, 61.
(368) Sharrocks, A. D. *Nat Rev Mol Cell Biol* **2001**, *2*, 827.
(369) Shakoori, A.; Ougolkov, A.; Yu, Z. W.; Zhang, B.; Modarressi, M. H.; Billadeau, D. D.; Mai, M.; Takahashi, Y.; Minamoto, T. *Biochem Biophys Res Commun* **2005**, *334*, 1365.
(370) Zhang, J. S.; Koenig, A.; Harrison, A.; Ugolkov, A. V.; Fernandez-Zapico, M. E.; Couch, F. J.; Billadeau, D. D. *Oncogene* **2011**, *30*, 3705.
(371) Ben-Shem, A.; Jenner, L.; Yusupova, G.; Yusupov, M. *Science* **2010**, *330*, 1203.

(372) Gregory, L. A.; Aguissa-Touré, A.-H.; Pinaud, N.; Legrand, P.; Gleizes, P.-E.; Fribourg, S. *Nucleic Acids Res* **2007**, *35*, 5913.
(373) Léger-Silvestre, I.; Caffrey, J. M.; Dawaliby, R.; Alvarez-Arias, D. A.; Gas, N.; Bertolone, S. J.; Gleizes, P.-E.; Ellis, S. R. *J Biol Chem* **2005**, *280*, 38177.
(374) Matsson, H.; Davey, E. J.; Draptchinskaia, N.; Hamaguchi, I.; Ooka, A.; Leveen, P.; Forsberg, E.; Karlsson, S.; Dahl, N. *Mol Cell Biol* **2004**, *24*, 4032.
(375) Choesmel, V.; Bacqueville, D.; Rouquette, J.; Noaillac-Depeyre, J.; Fribourg, S.; Cretien, A.; Leblanc, T.; Tchernia, G.; Da Costa, L.; Gleizes, P.-E. *Blood* **2007**, *109*, 1275.
(376) Flygare, J.; Aspesi, A.; Bailey, J. C.; Miyake, K.; Caffrey, J. M.; Karlsson, S.; Ellis, S. R. *Blood* **2007**, *109*, 980.
(377) Idol, R. A.; Robledo, S.; Du, H.-Y.; Crimmins, D. L.; Wilson, D. B.; Ladenson, J. H.; Bessler, M.; Mason, P. J. *Blood Cell Mol Dis* **2007**, *39*, 35.
(378) Gazda, H. T.; Grabowska, A.; Merida-Long, L. B.; Latawiec, E.; Schneider, H. E.; Lipton, J. M.; Vlachos, A.; Atsidaftos, E.; Ball, S. E.; Orfali, K. A.; Niewiadomska, E.; Da Costa, L.; Tchernia, G.; Niemeyer, C.; Meerpohl, J. J.; Stahl, J.; Schratt, G.; Glader, B.; Backer, K.; Wong, C.; Nathan, D. G.; Beggs, A. H.; Sieff, C. A. *The American Journal of Human Genetics* **2006**, *79*, 1110.
(379) Cmejla, R.; Cmejlova, J.; Handrkova, H.; Petrak, J.; Pospisilova, D. *Hum Mutat* **2007**, *28*, 1178.
(380) Ellis, S. R.; Massey, A. T. *Med Hypotheses* **2006**, *66*, 643.
(381) Willig, T. b.-N. l.; Gazda, H.; Sieff, C. A. *Curr Opin Hematol* **2000**, *7*, 85.
(382) Costa, L. D.; Willig, T.-N.; Fixler, J.; Mohandas, N.; Tchernia, G. *Curr Opin Pediatr* **2001**, *13*, 10.
(383) Kondoh, N.; Schweinfest, C. W.; Henderson, K. W.; Papas, T. S. *Cancer Res* **1992**, *52*, 791.
(384) Bachmeier, B.; Fichtner, I.; Killian, P. H.; Kronski, E.; Pfeffer, U.; Efferth, T. *PLoS One* **2011**, *6*, e20550.
(385) Adams, J. M.; Cory, S. *Oncogene* **2007**, *26*, 1324.
(386) Willis, S. N.; Adams, J. M. *Curr Opin Cell Biol* **2005**, *17*, 617.
(387) Huang, D. C. S.; Strasser, A. *Cell* **2000**, *103*, 839.
(388) Youle, R. J.; Strasser, A. *Nat Rev Mol Cell Biol* **2008**, *9*, 47.
(389) Costantino, V.; Fattorusso, E.; Mangoni, A. *J Org Chem* **1993**, *58*, 186.
(390) Cowgill, U. M.; Milazzo, D. P. *The Culturing and Testing of Two Species of Duckweed.*; American Society for Testing and Material International: Philadelphia, 1989; Vol. 12.
(391) Spura, A.; Riel, R. U.; Freedman, N. D.; Agrawal, S.; Seto, C.; Hawrot, E. *J Biol Chem* **2000**, *275*, 22452.
(392) Ansari, M.; Sunderland, V. *Arch Pharmacal Res* **2008**, *31*, 390.
(393) Haynes, R. K.; Chan, H. W.; Lung, C. M.; Ng, N. C.; Wong, H. N.; Shek, L. Y.; Williams, I. D.; Cartwright, A.; Gomes, M. F. *ChemMedChem* **2007**, *2*, 1448.
(394) Novagen In *User Protocol TB178*; Rev. D 0311JN ed. USA, 2011, p 1.
(395) Weinzierl, A. O.; Stevanovic, S. *Biotechnol Genet Eng* **2006**, *23*, 21.
(396) Eberl, H. C.; Mann, M.; Vermeulen, M. *ChemBioChem* **2011**, *12*, 224.
(397) Invitrogen 2010; Vol. Manual part no. 28-0182, p 20.
(398) Schleucher, J.; Quant, J.; Glaser, S. J.; Griesinger, C. In *Encyclopedia of Magnetic Resonance*; John Wiley & Sons, Ltd: 2007, p 259.
(399) Rozen, S.; Skaletsky, H. J. In *Bioinformatics Methods and Protocols: Methods in Molecular Biology*; Krawetz, S., Misener, S., Eds.; Humana Press: Totowa, 2000, p 365.

Appendices

8.1 Marine natural products

8.1.1 Pseudoceratina purpurea



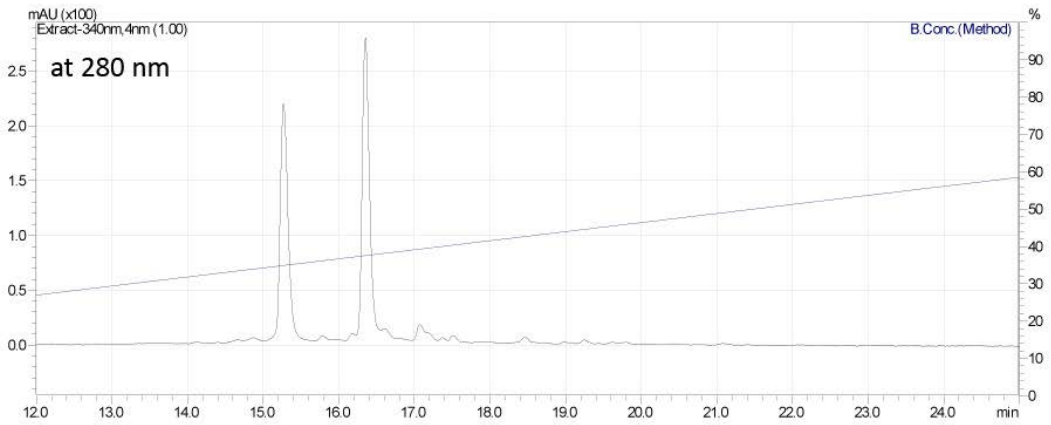
8.1.1.1 Bioassay guided fractionation

8.1.1.1.1 Gel permeation filtration of *P. purpurea* ethyl acetate partition

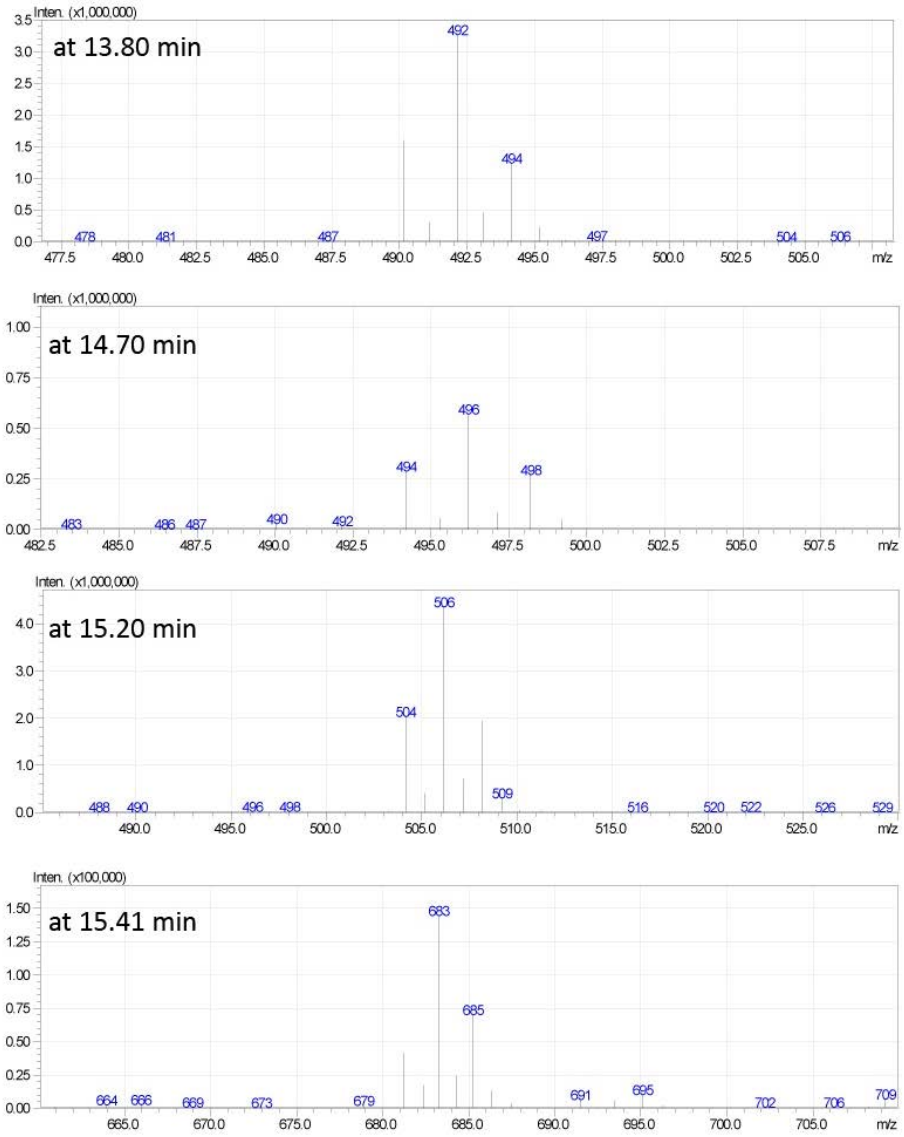
P. purpurea **Sephadex fraction**	**Dry Weight (in mg)**	**Disc Diffusion Inhibition Zone (in mm with 300 µg)**		
		E. coli	*S. aureus*	*P. aeruginosa*
1-8	197.1	< 7	< 7	< 7
9-12	720.1	14	14	< 7
13-14	537.7	16	16	< 7
15-22	403.6	10	10	< 7
23-25	46.4	< 7	< 7	< 7
26-27	2.6	< 7	< 7	< 7
28	1.8	< 7	< 7	< 7
29-30	4.0	< 7	< 7	< 7
31	2.5	< 7	< 7	< 7
32-37	41.8	< 7	< 7	< 7
39-44	8.3	< 7	< 7	< 7
45-52	8.6	< 7	< 7	< 7
53-60	25.3	< 7	< 7	< 7
61-71	19.6	< 7	< 7	< 7
75	4.6	< 7	< 7	< 7
72-86	50.6	< 7	< 7	< 7
87	34.5	< 7	< 7	< 7
88-90	16.1	< 7	< 7	< 7
total	2125.2	n.a.		

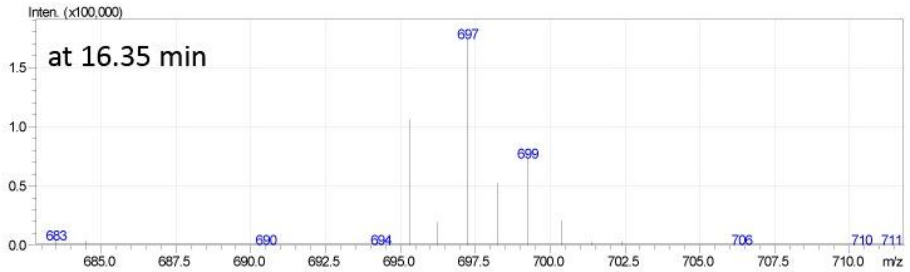
8.1.2 UV trace and MS data of ethyl acetate fractions of *P. purpurea* and *T. corticalis*

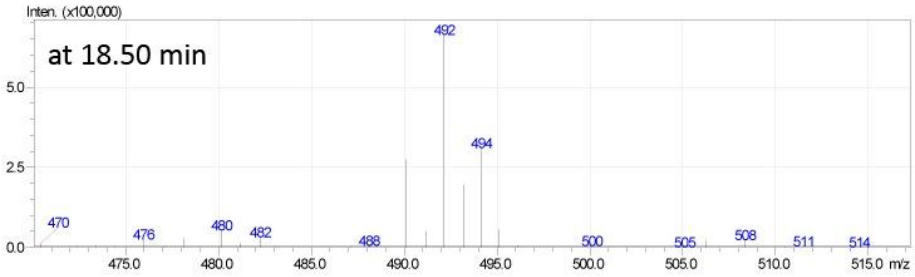
8.1.2.1 UV trace of the ethyl acetate fraction of P. purpurea at 280 nm

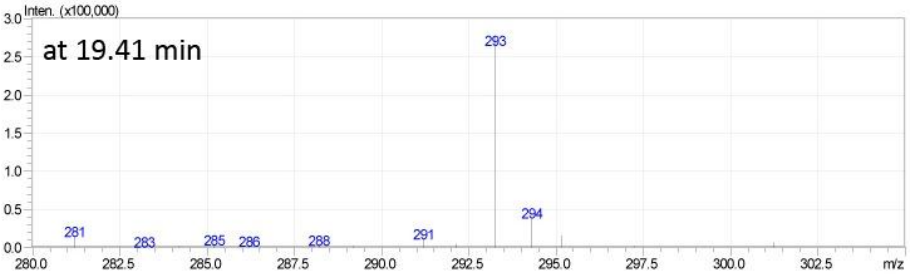


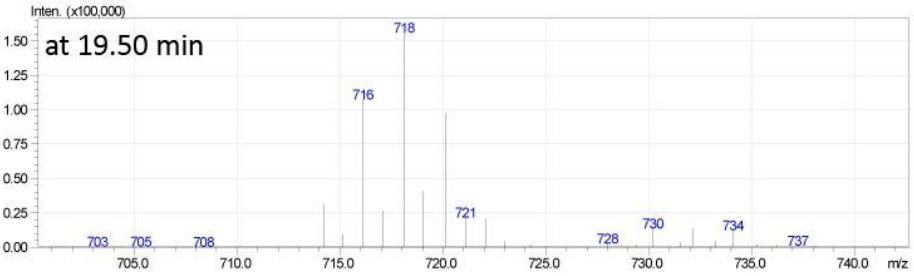
8.1.2.2 Selective ion traces of compounds listed in Table 11

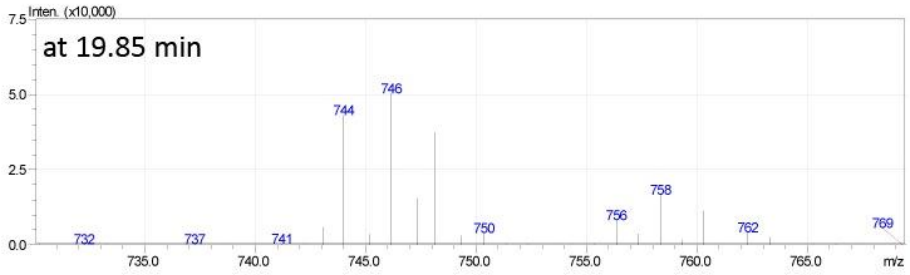


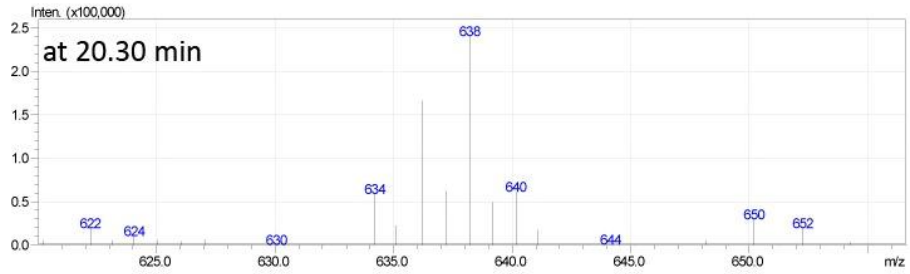

Inten. (x100,000)
at 16.35 min
697
699
683
690
694
706
710
711
m/z

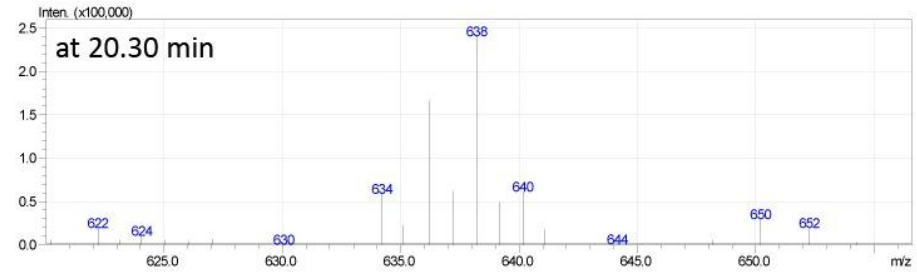

Inten. (x100,000)
at 18.50 min
492
494
470
476
480
482
488
500
505
508
511
514
m/z

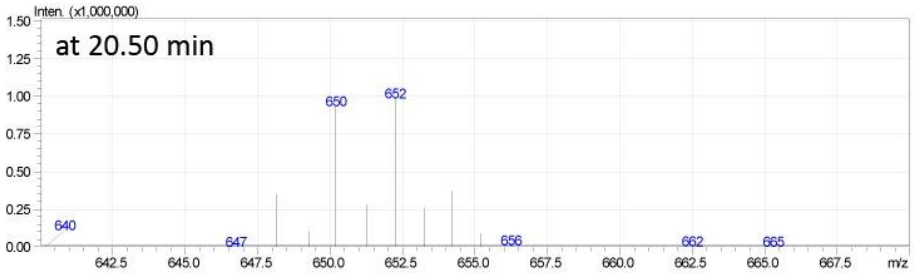

Inten. (x100,000)
at 19.41 min
293
294
281
283
285
286
288
291
m/z

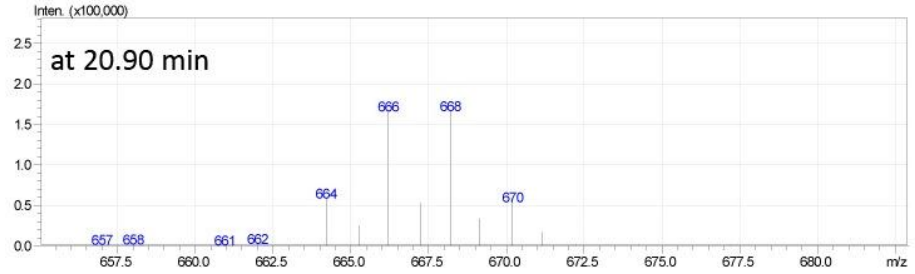

Inten. (x100,000)
at 19.50 min
718
716
721
703
705
708
728
730
734
737
m/z

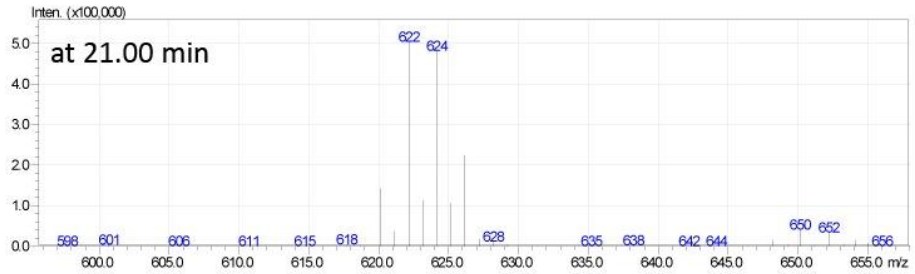

Inten. (x10,000)
at 19.85 min
746
744
758
732
737
741
750
756
762
769
m/z

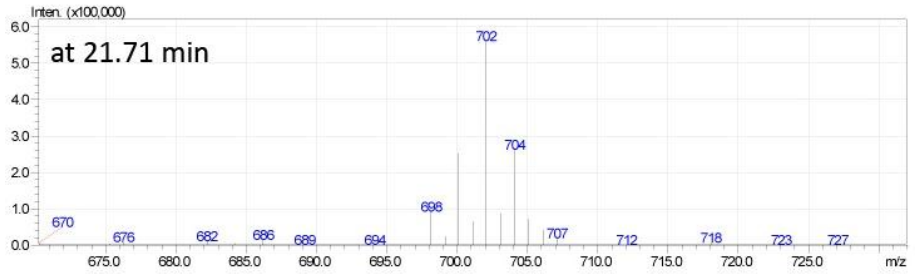

Inten. (x100,000)
at 20.30 min
638
634
640
622
624
630
644
650
652
m/z

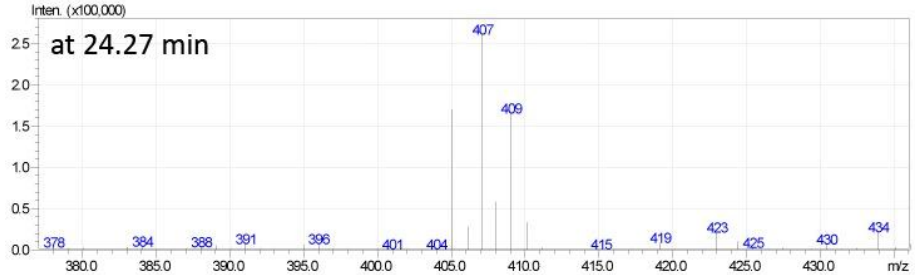

Inten. (x100,000)
at 20.30 min
638
634
640
622
624
630
644
650
652
m/z


Inten. (x1,000,000)
at 20.50 min
650
652
640
647
656
662
665
m/z


Inten. (x100,000)
at 20.90 min
666
668
664
670
657
658
661
662
m/z

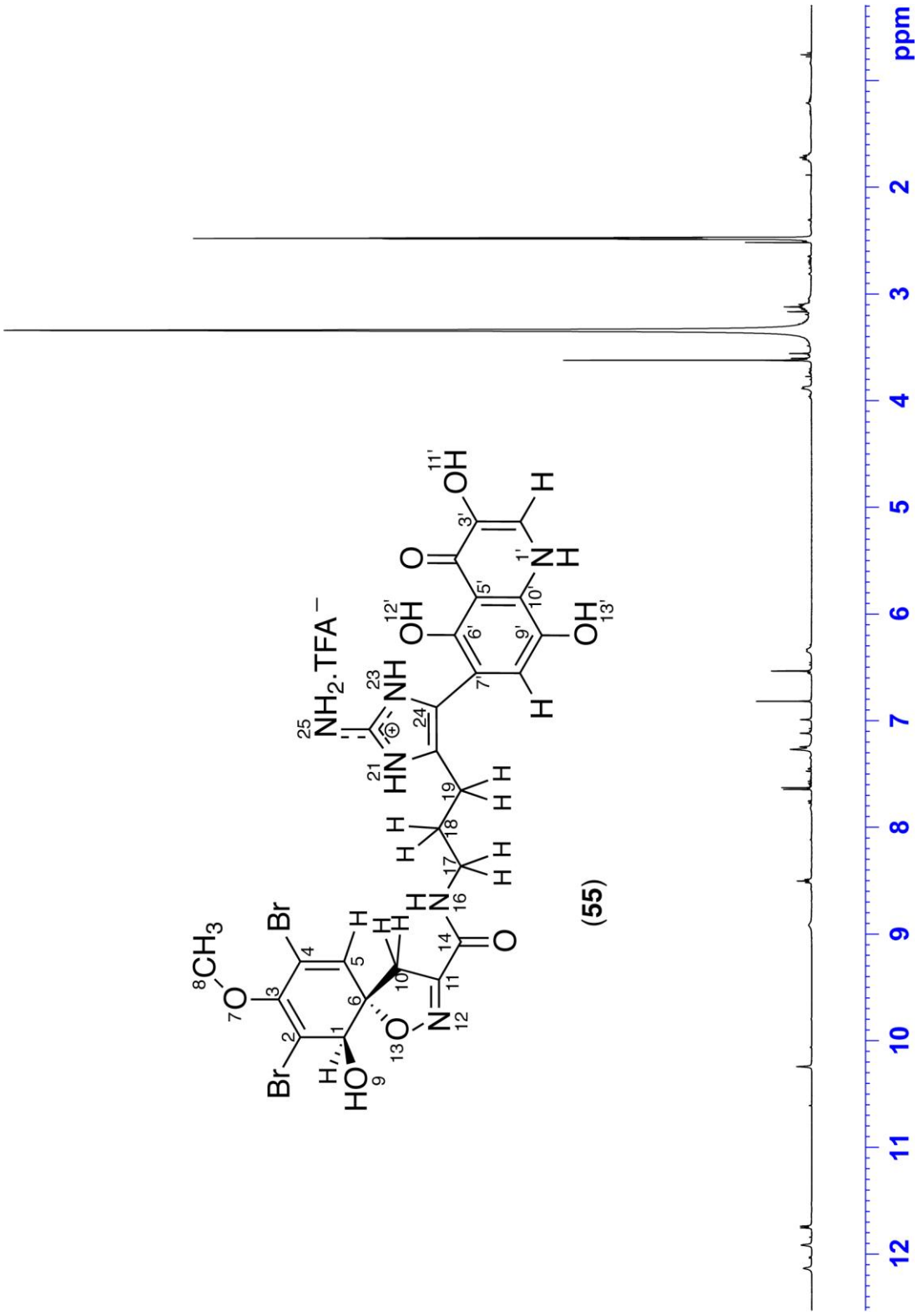

Inten. (x100,000)
at 21.00 min
622
624
598
601
606
611
615
618
628
635
638
642
644
650
652
656
m/z


Inten. (x100,000)
at 21.71 min
702
704
698
670
676
682
686
689
694
707
712
718
723
727
m/z

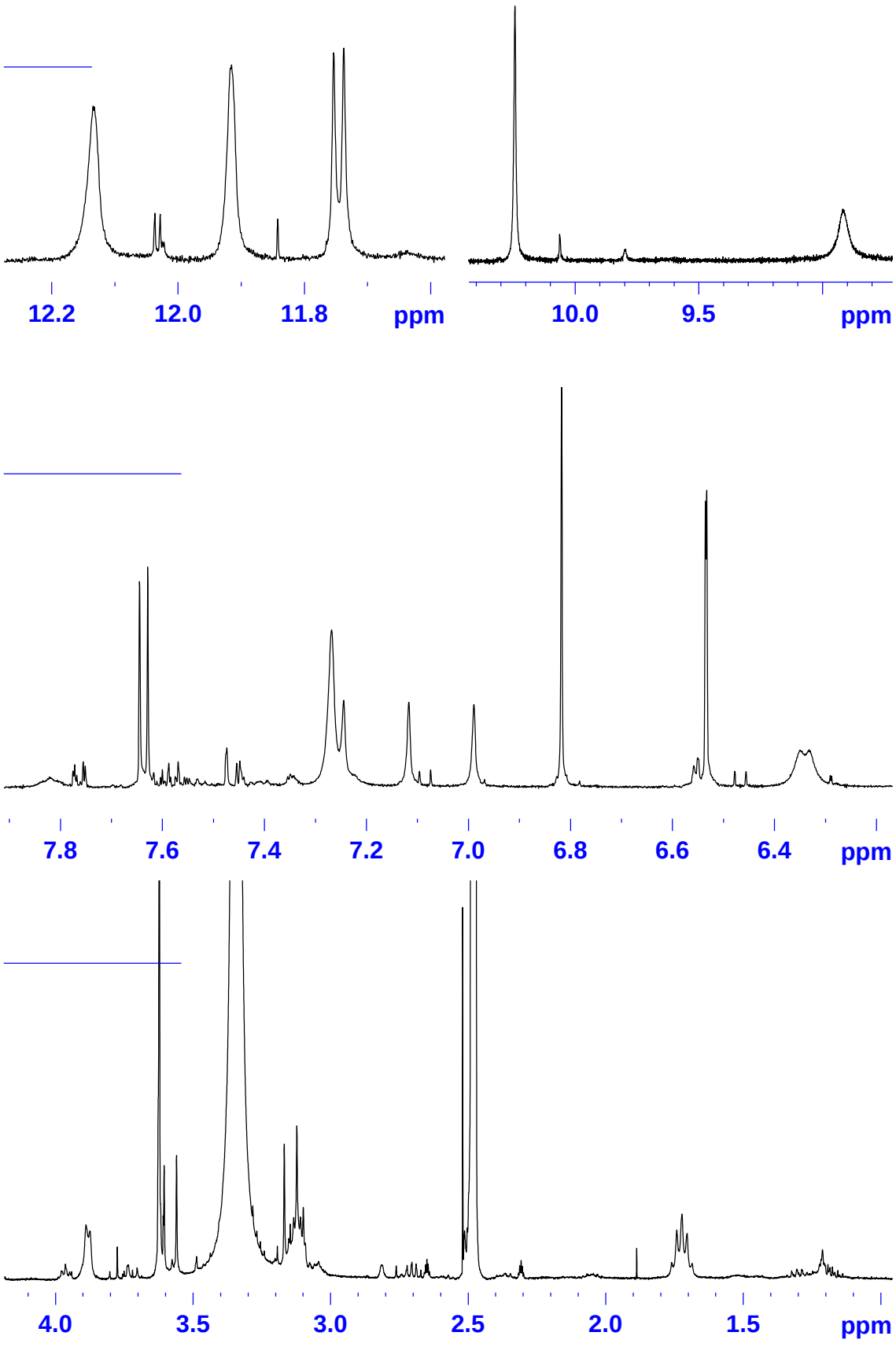

Inten. (x100,000)
at 24.27 min
407
409
378
384
388
391
396
401
404
415
419
423
425
430
434
m/z

8.1.3 NMR spectra of bromotyrosine (55; ceratinadin D)

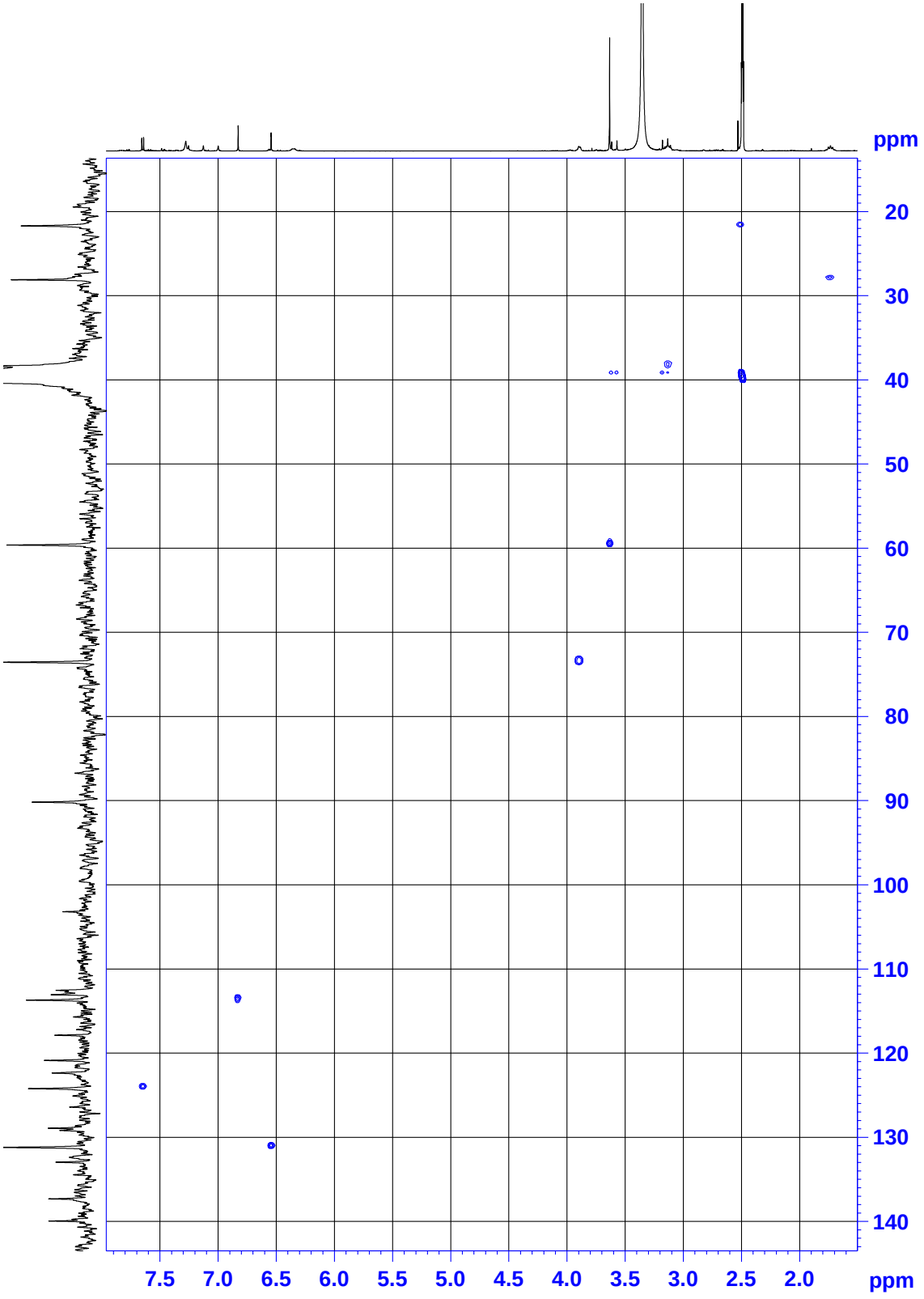
8.1.3.1 ^{1}H NMR spectrum of (55; DMSO-d_6, 600 MHz)



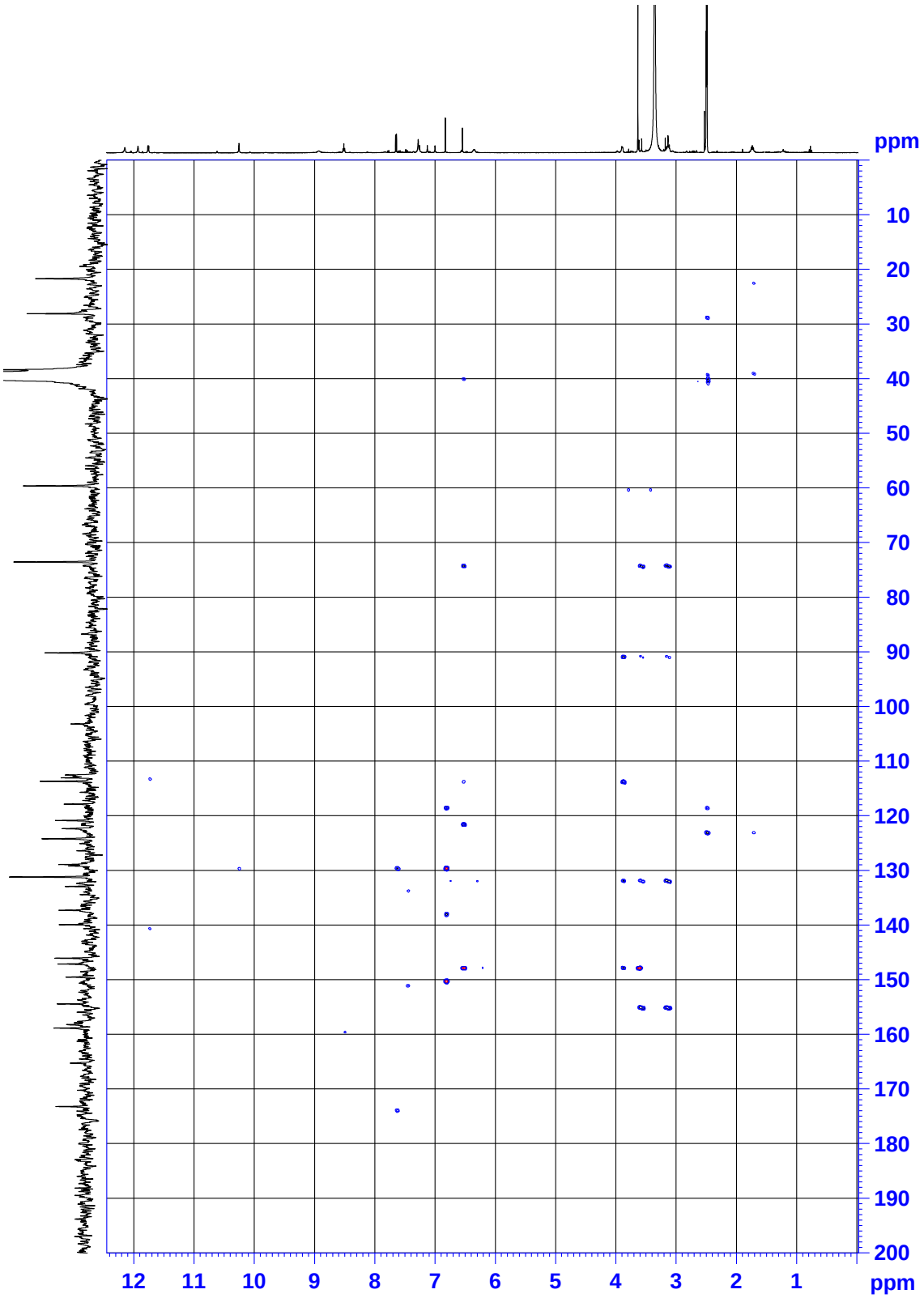
8.1.3.2 ^{1}H NMR spectrum of (55; DMSO-d_6, 600 MHz) - expansions



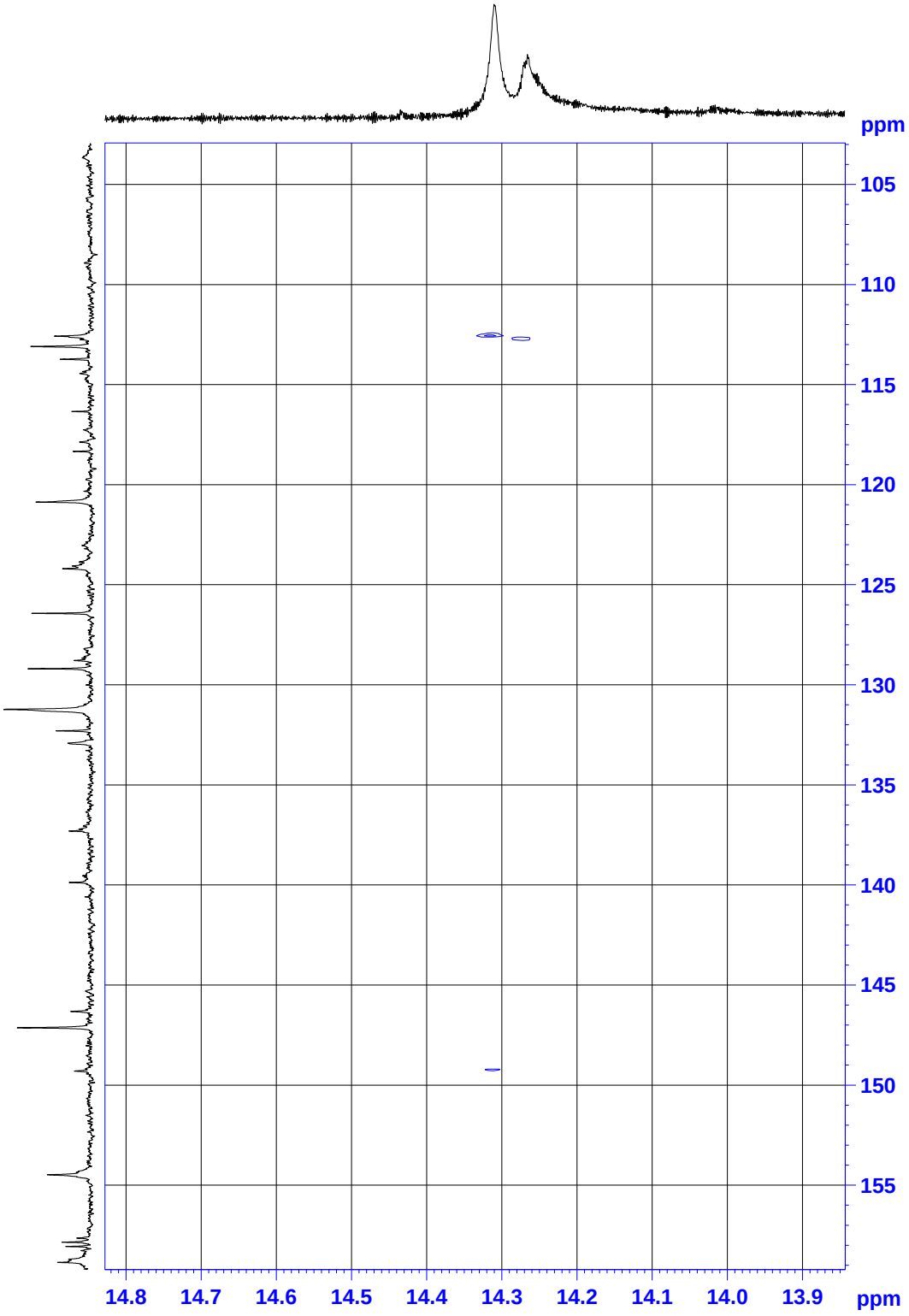
8.1.3.3 HSQC spectrum of (55; DMSO-d_6, 600 MHz)



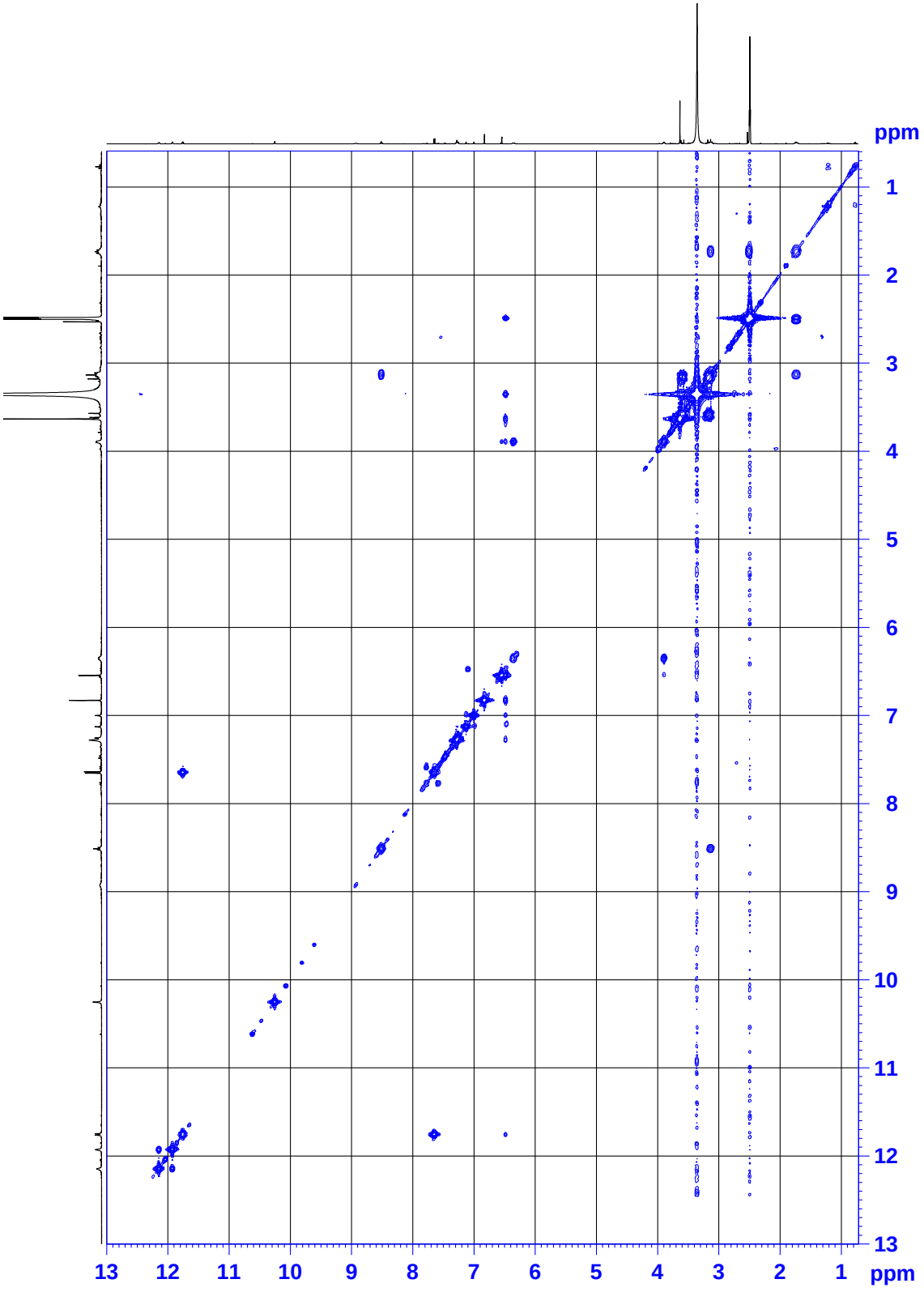
8.1.3.4 HMBC spectrum of (55; DMSO-d_6, 600 MHz)



8.1.3.5 HMBC spectrum of (55; DMSO-d_6, 600 MHz) – expansions

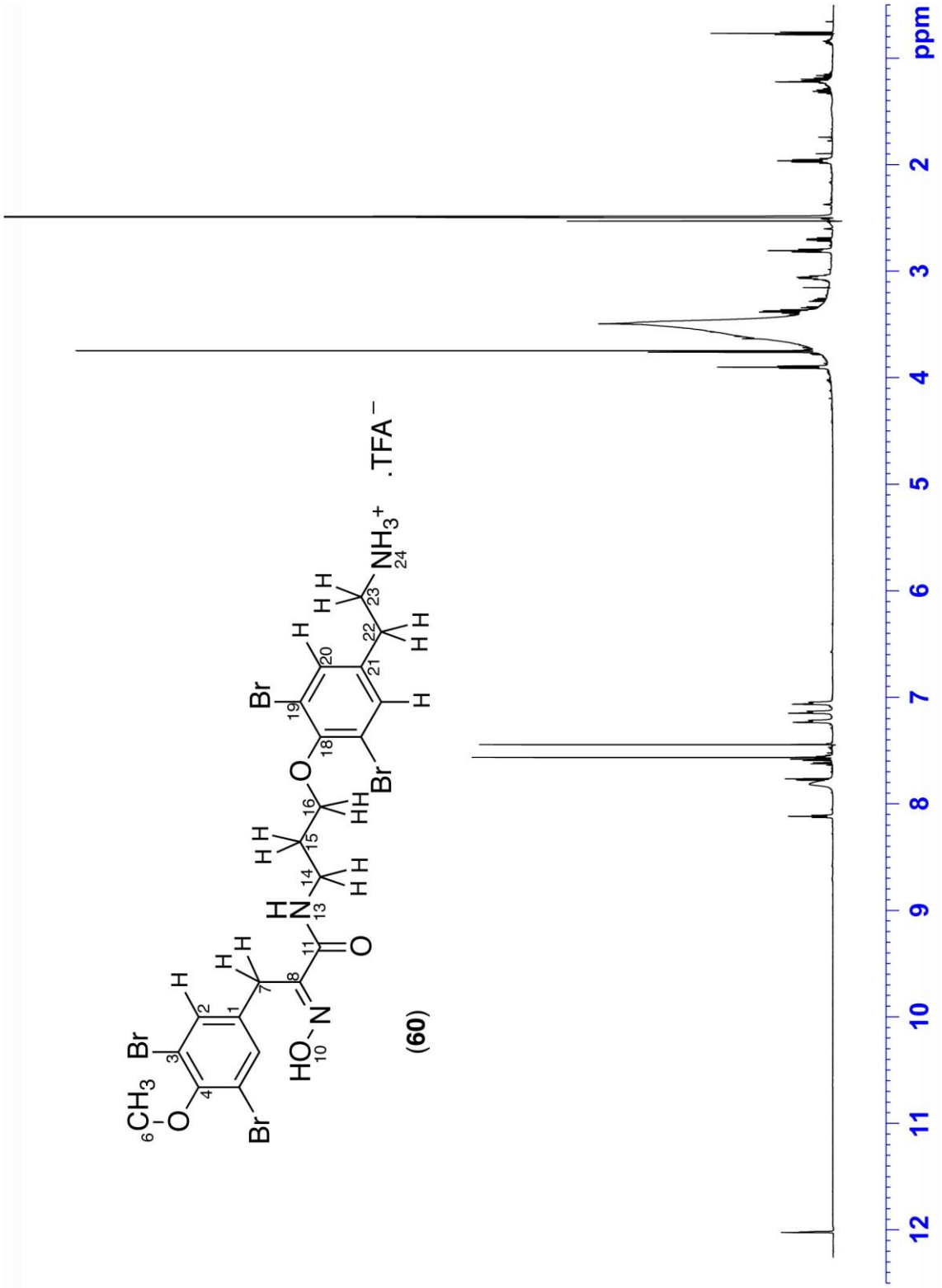


8.1.3.6 COSY spectrum of (55; DMSO-d_6, 600 MHz)

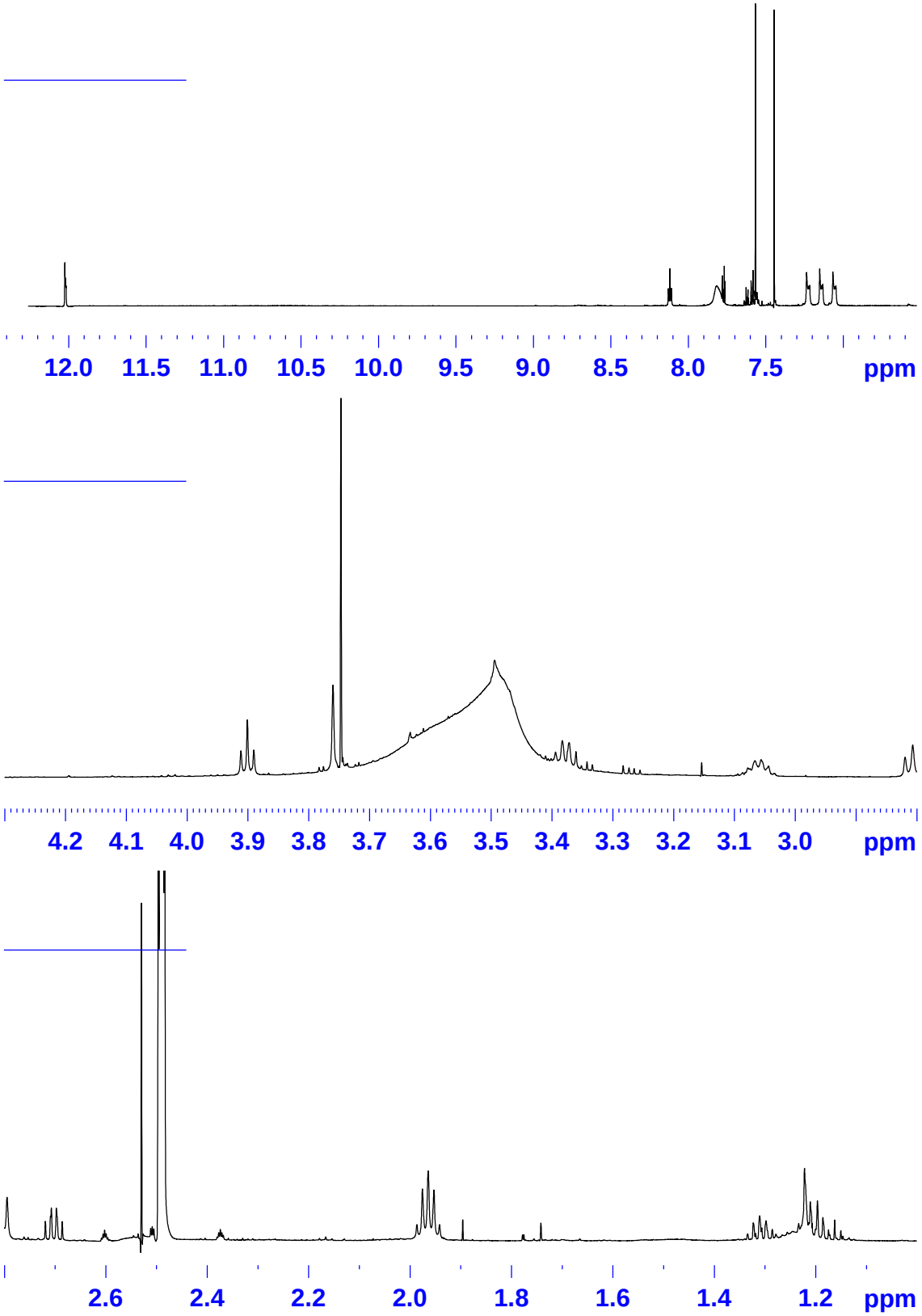


8.1.4 NMR spectra of bromotyrosine (60)

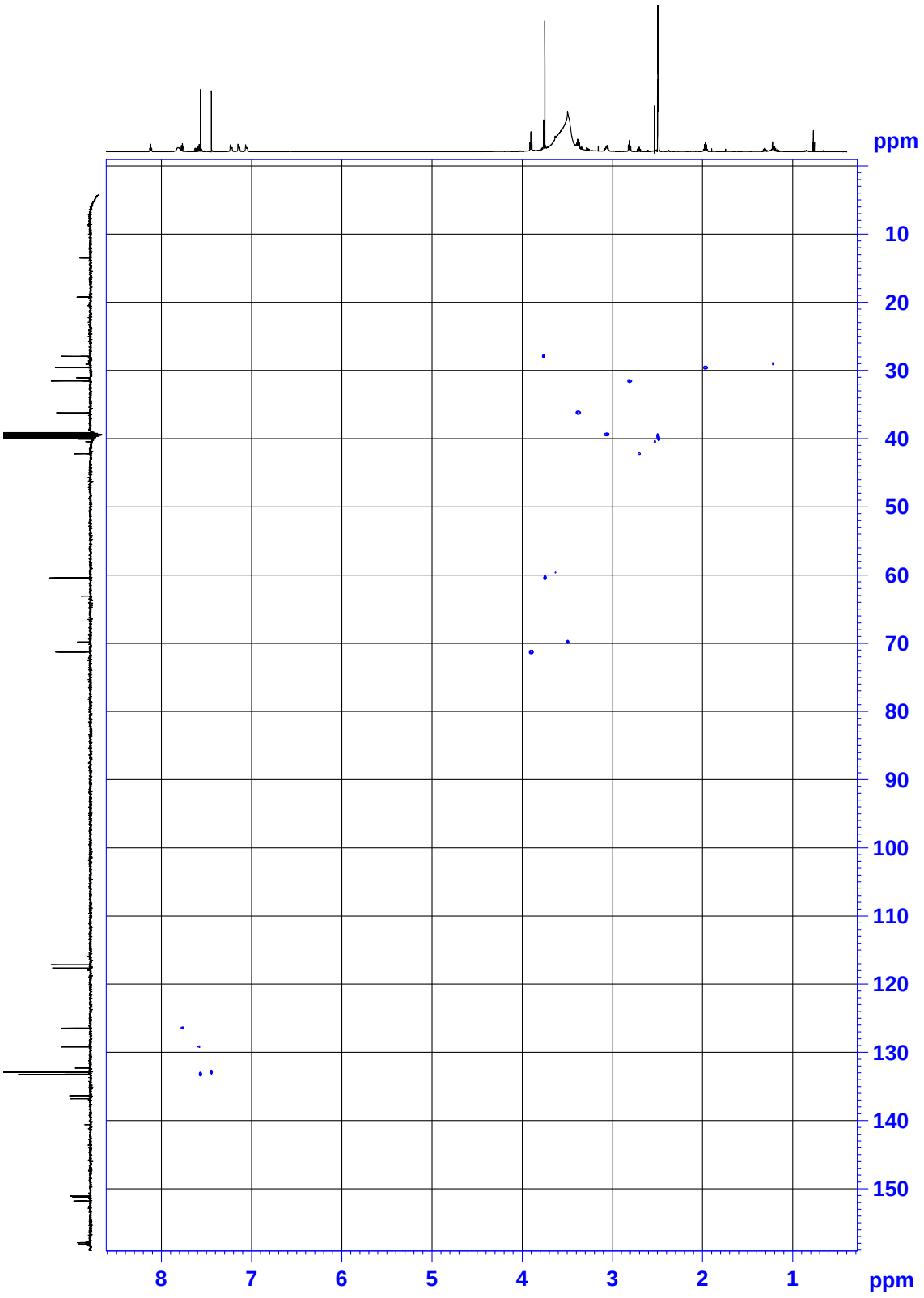
8.1.4.1 ^{1}H NMR spectrum of (60; DMSO-d_6, 600 MHz)



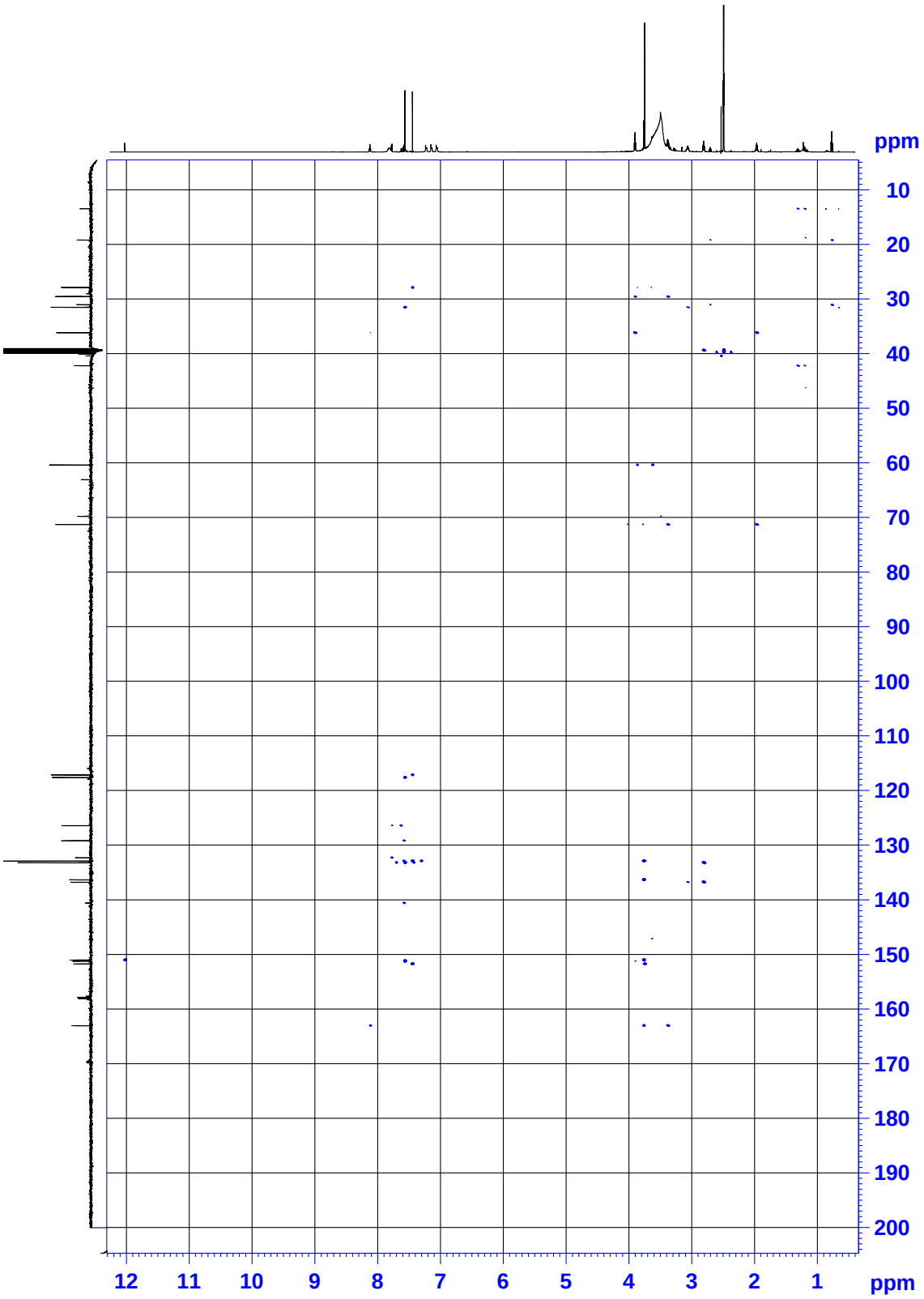
8.1.4.2 ^{1}H NMR spectrum of (60; DMSO-d_6, 600 MHz) – expansions



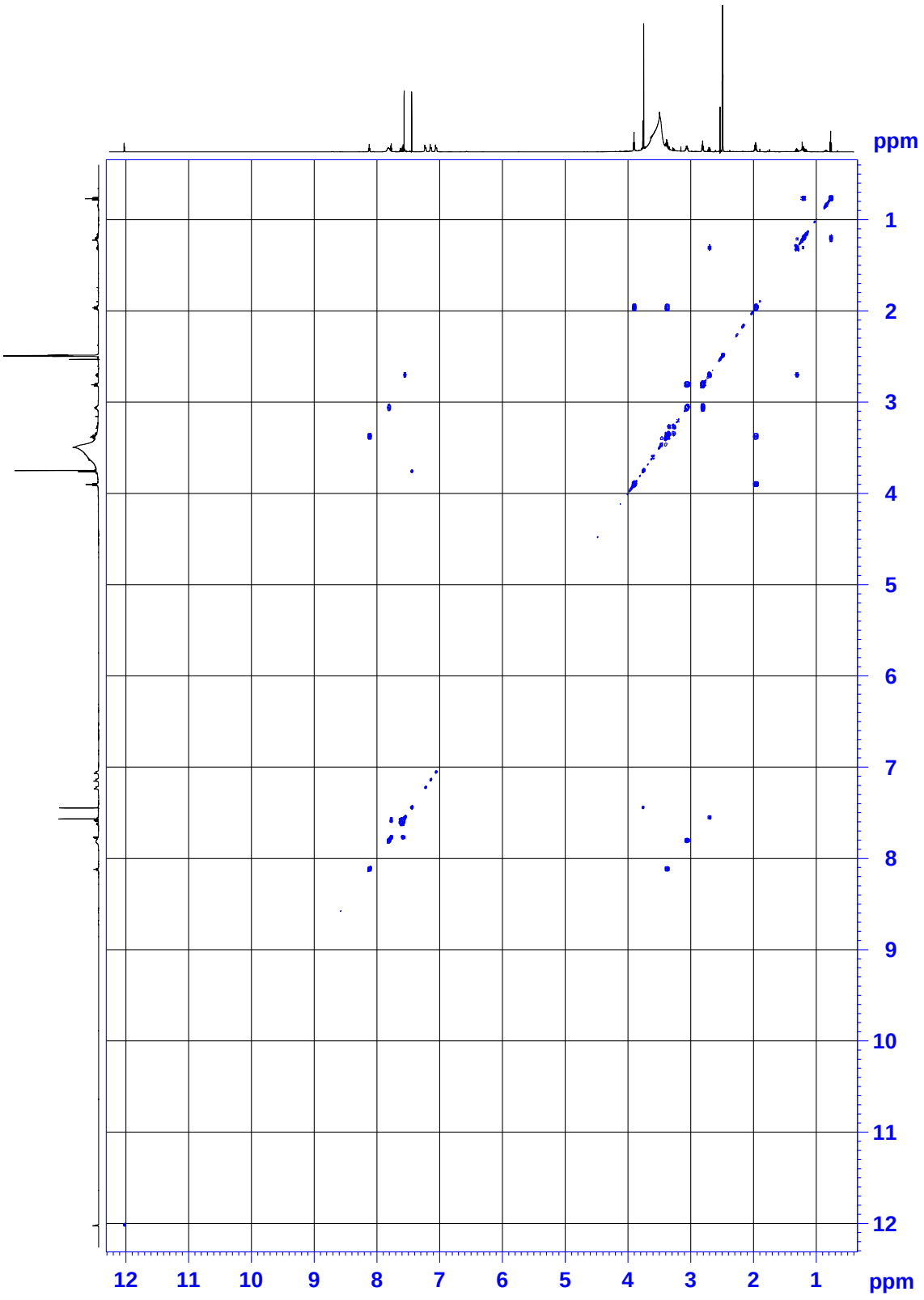
8.1.4.3 HSQC spectrum of (60; DMSO-d_6, 600 MHz)



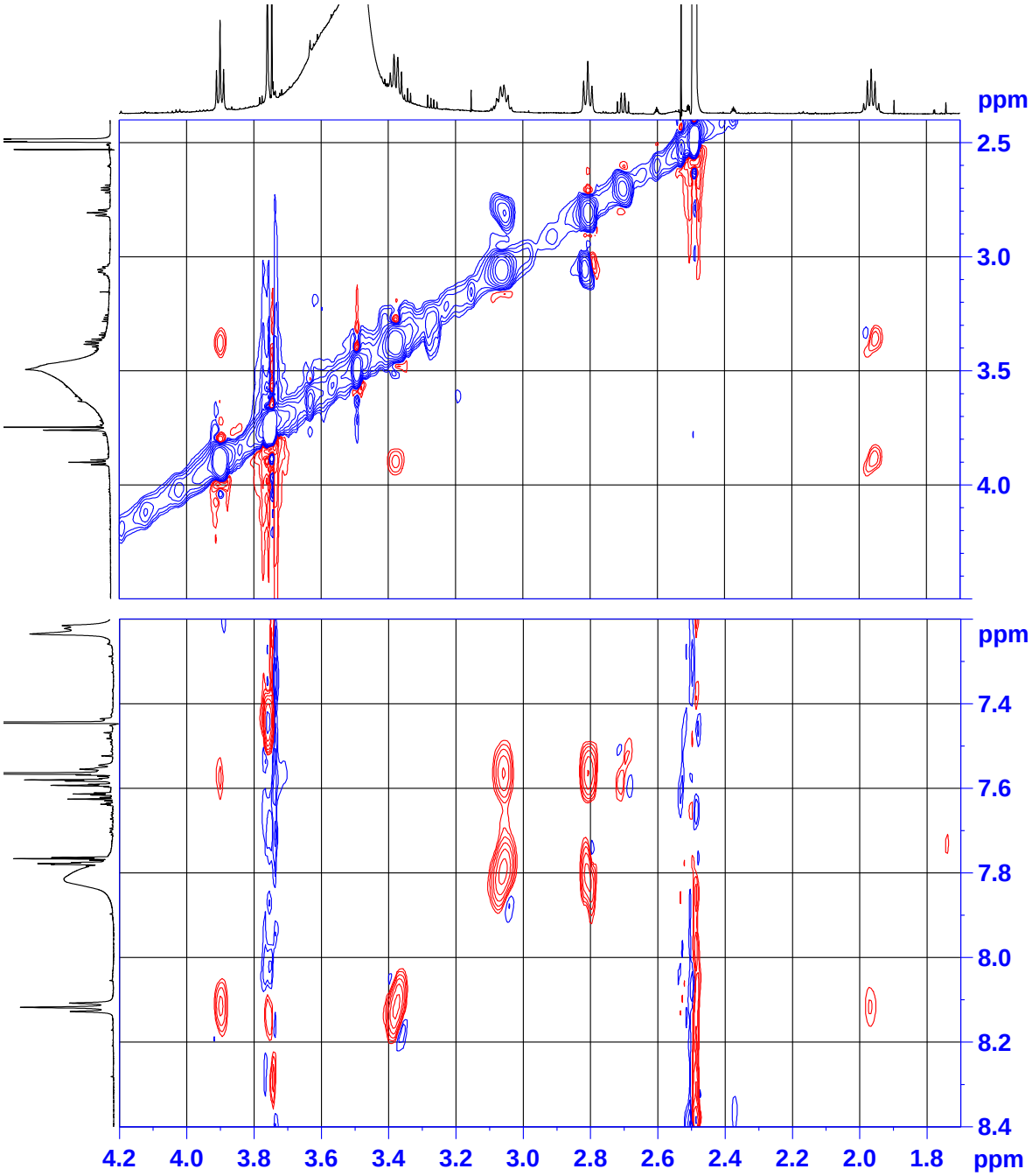
8.1.4.4 HMBC spectrum of (60; DMSO-d_6, 600 MHz)



8.1.4.5 COSY spectrum of (60; DMSO-d_6, 600 MHz)



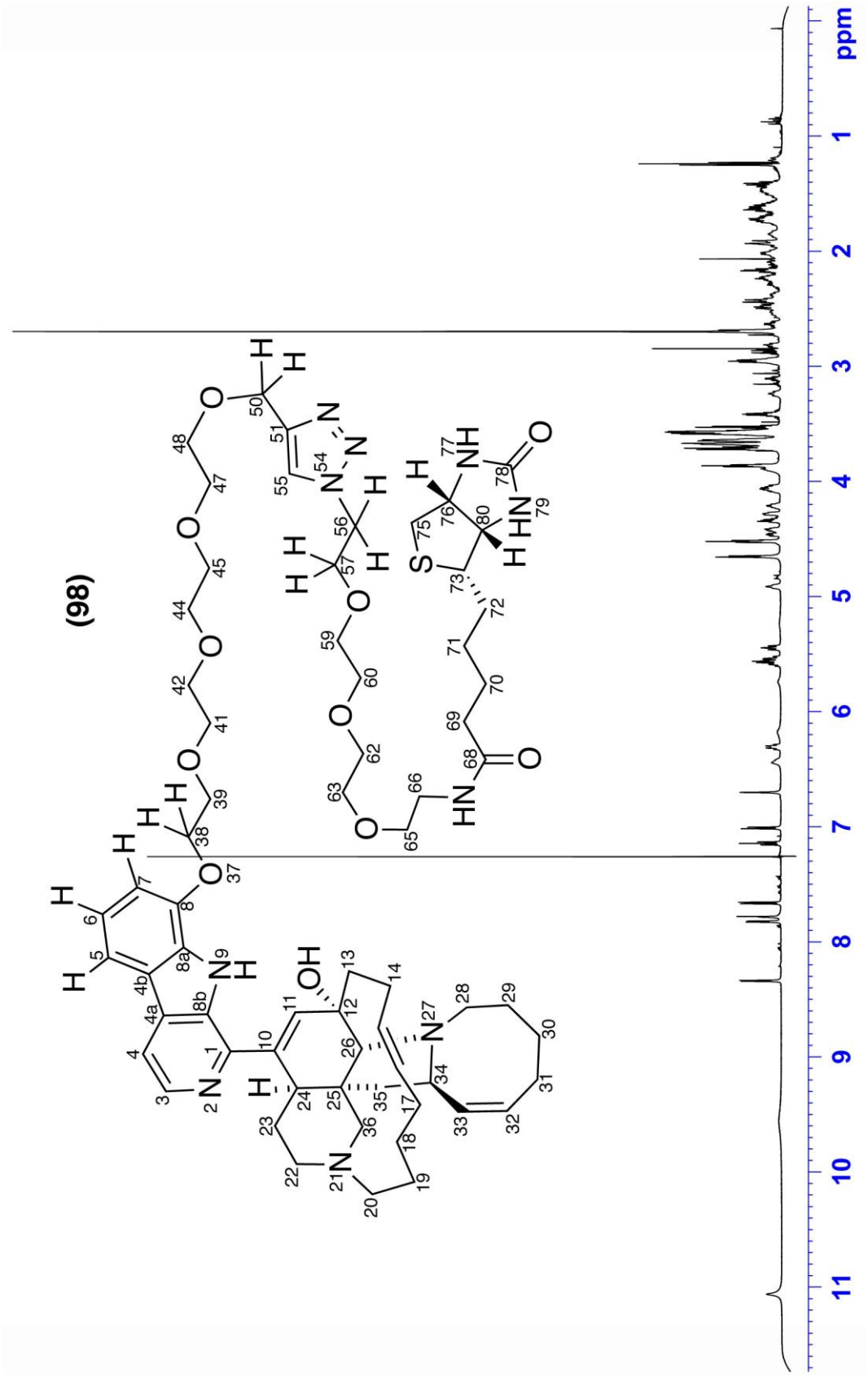
8.1.4.6 ROESY spectrum (100 ms) of (60; DMSO-d_6, 600 MHz)



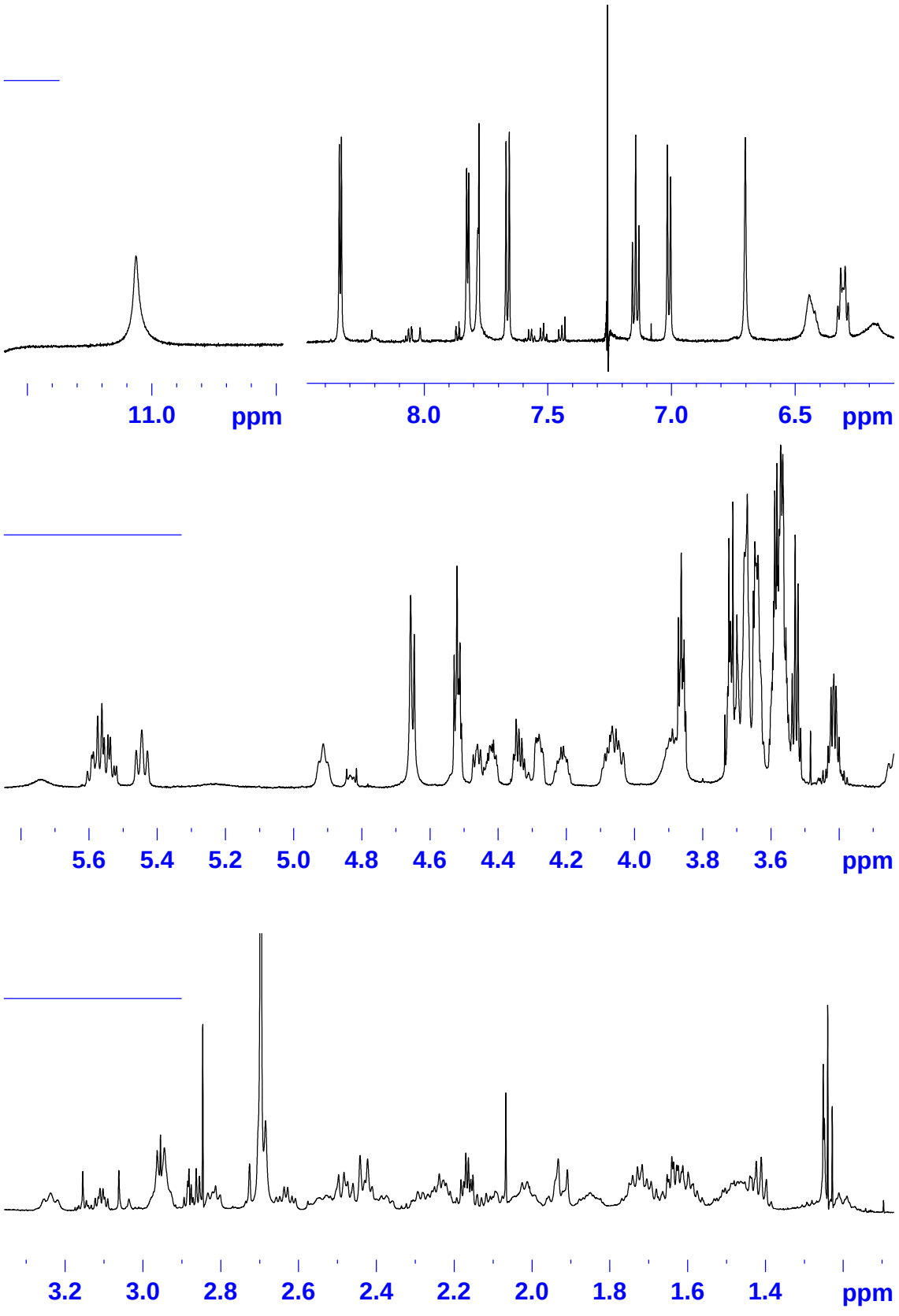
8.2 Synthesis of reagents and linkers

8.2.1 NMR spectra of biotin-manzamine (98)

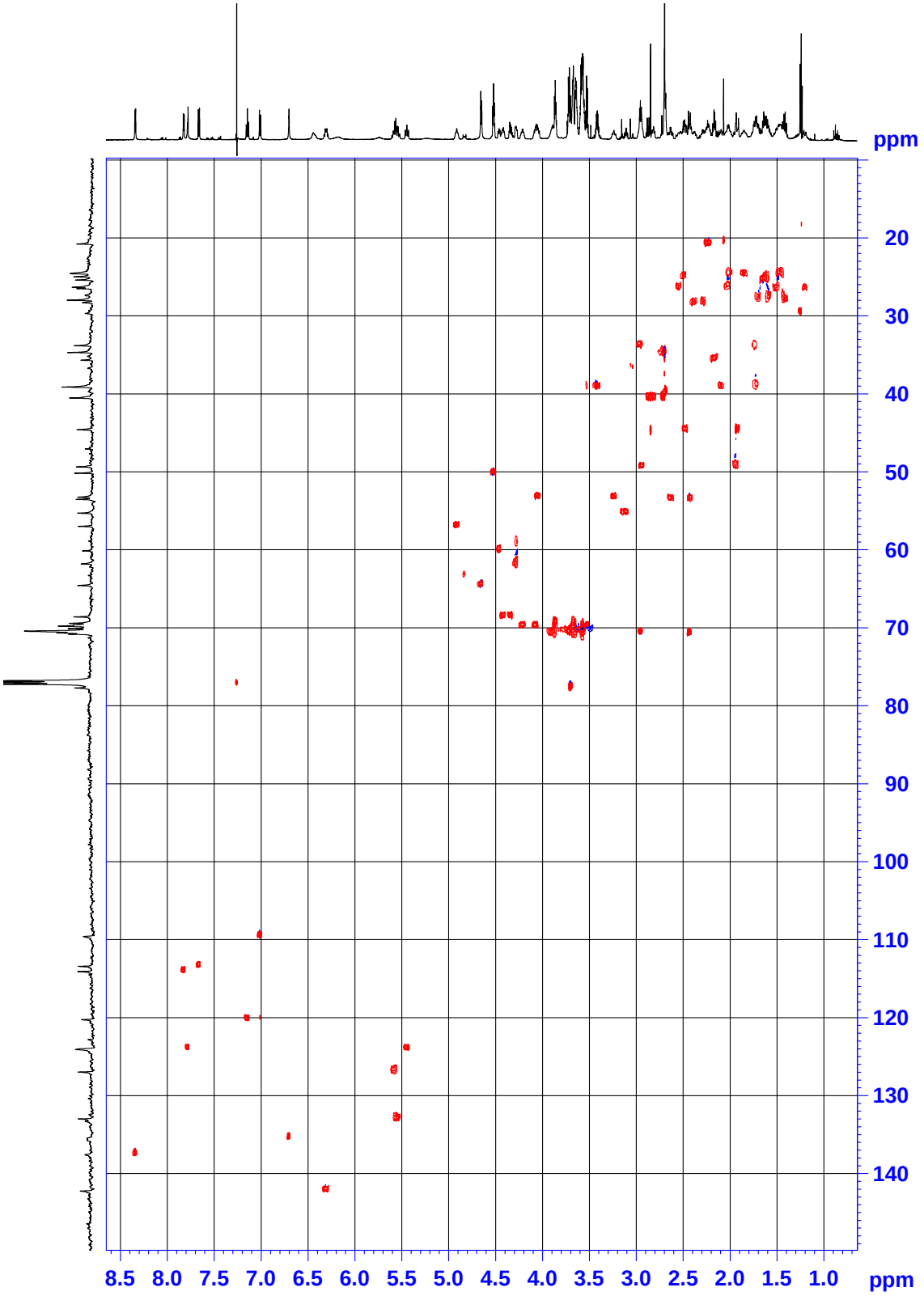
8.2.1.1 ^{1}H NMR spectrum of biotin-manzamine (98; $CDCl_3$, 600 MHz)



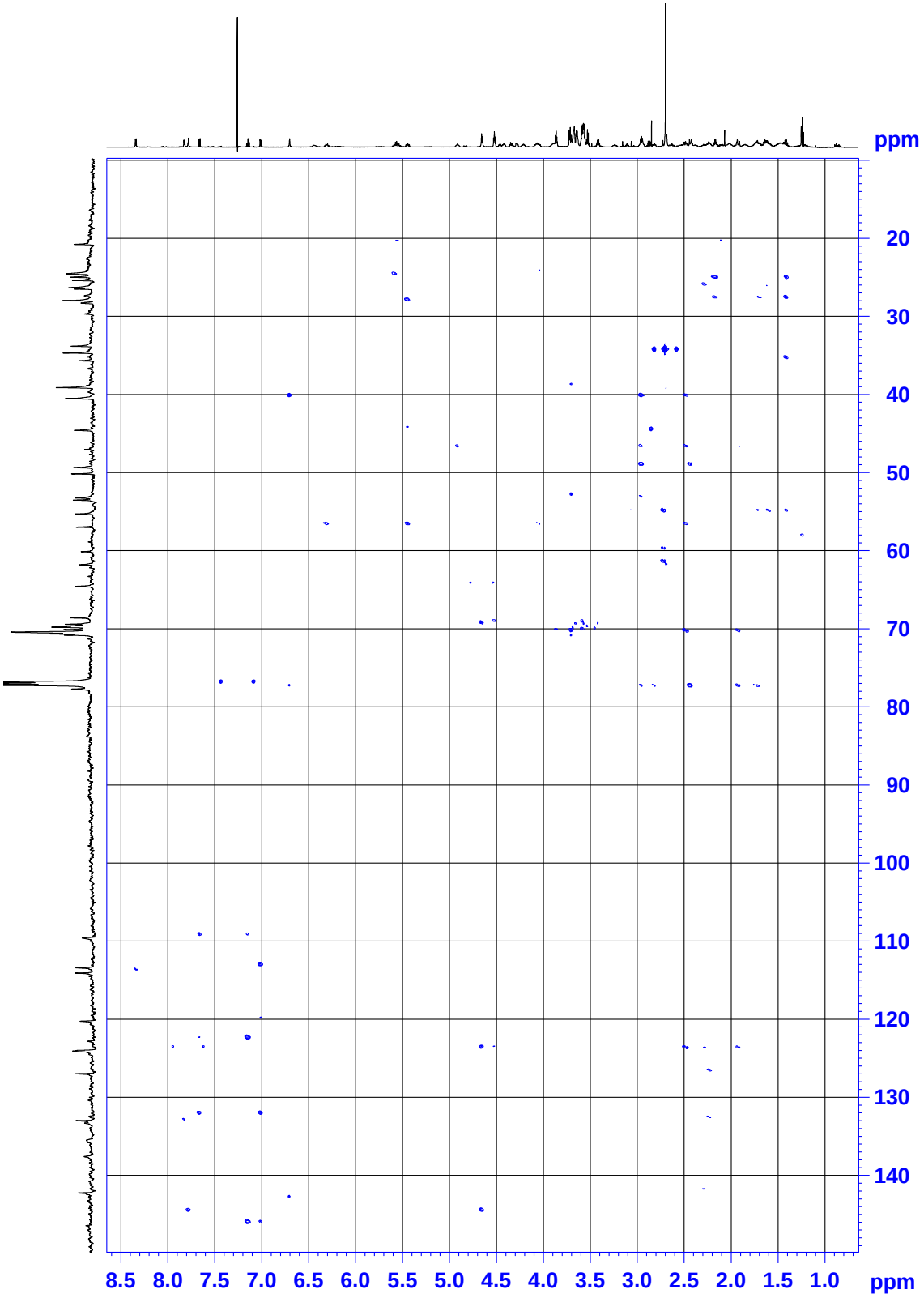
8.2.1.2 ^{1}H NMR spectrum of biotin-manzamine (98; $CDCl_3$, 600 MHz) – expansions



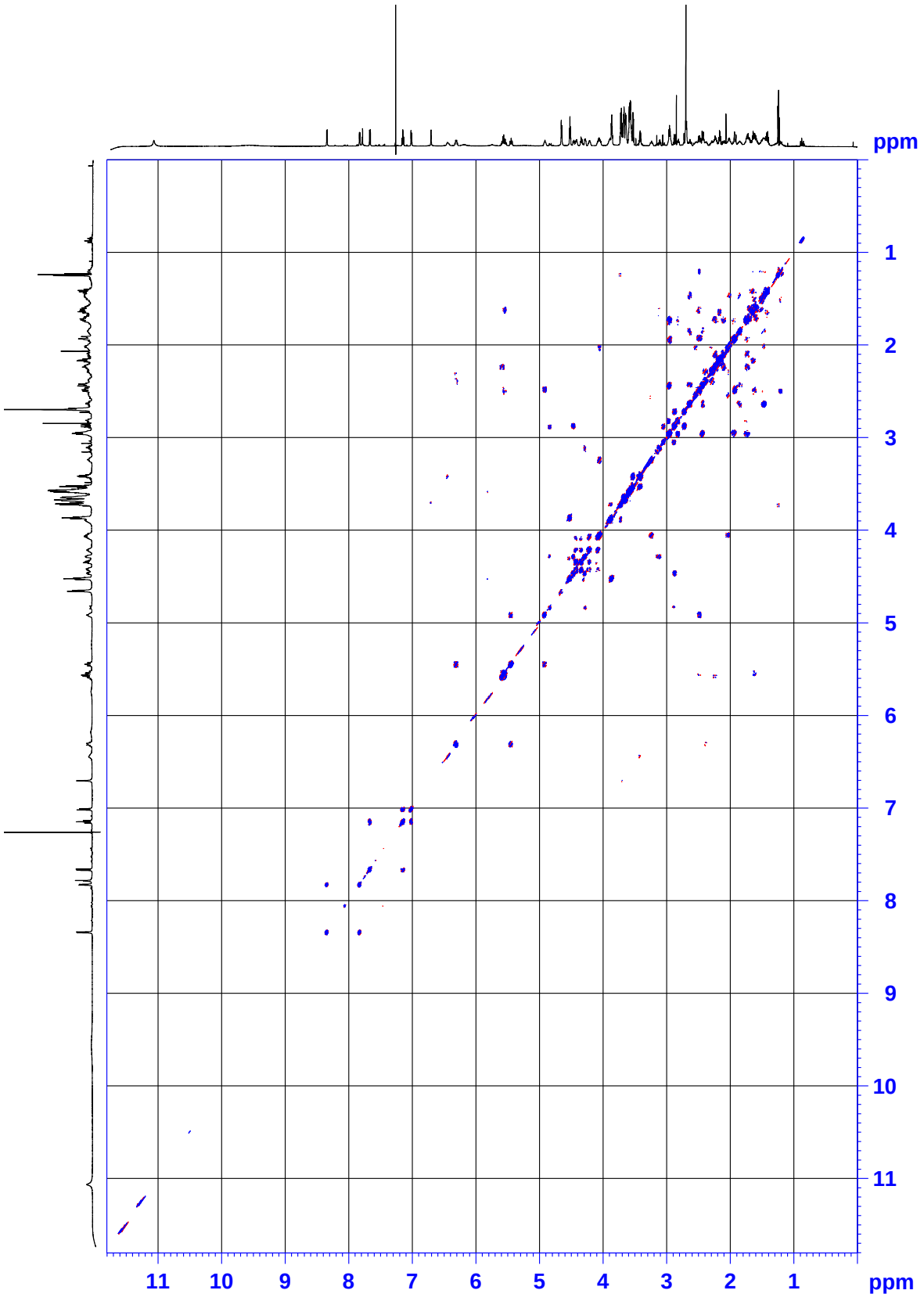
8.2.1.3 HSQC spectrum of biotin-manzamine (98; $CDCl_3$, 600 MHz)



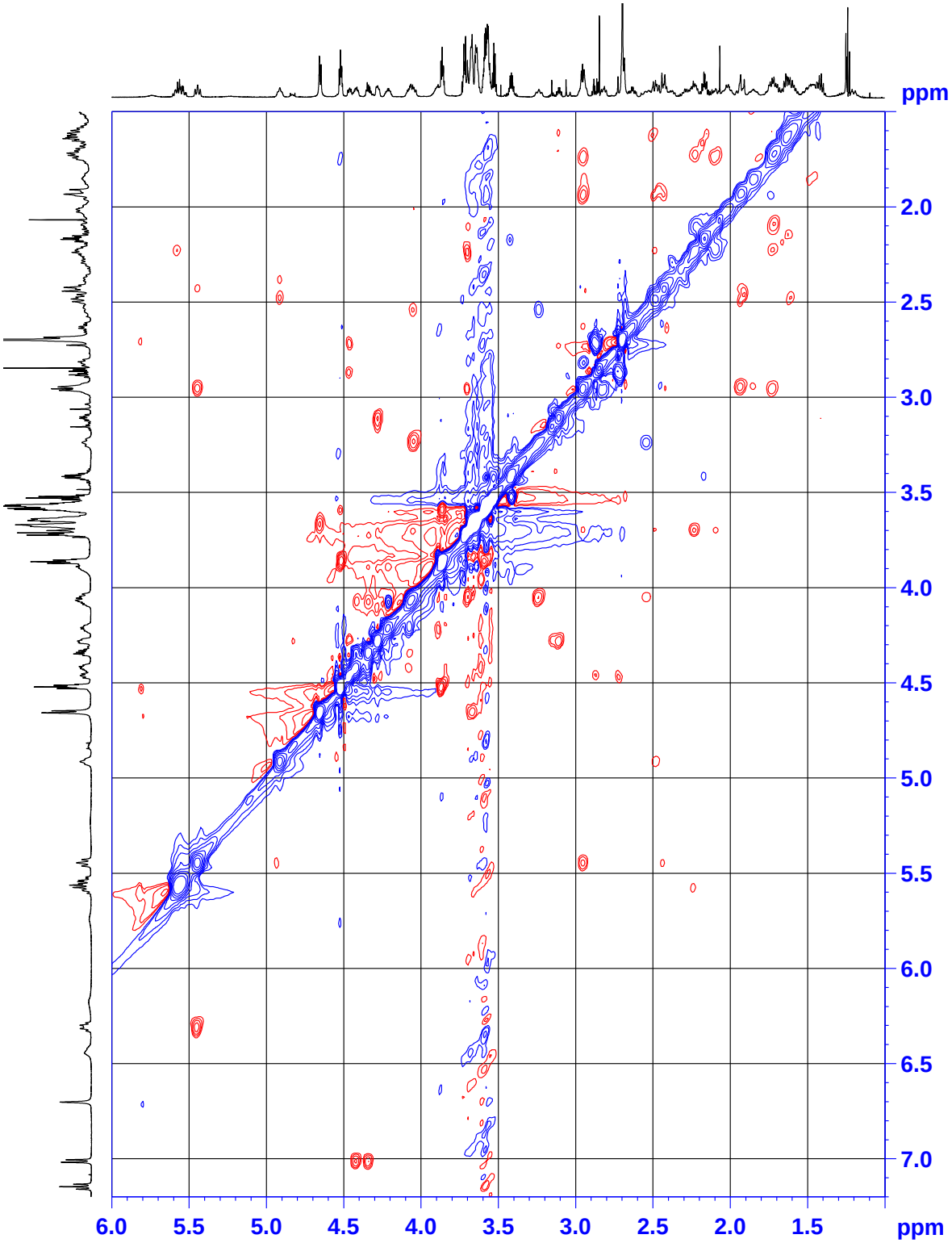
8.2.1.4 HMBC spectrum of biotin-manzamine (98; $CDCl_3$, 600 MHz)



8.2.1.5 COSY spectrum of biotin-manzamine (98; $CDCl_3$, 600 MHz)

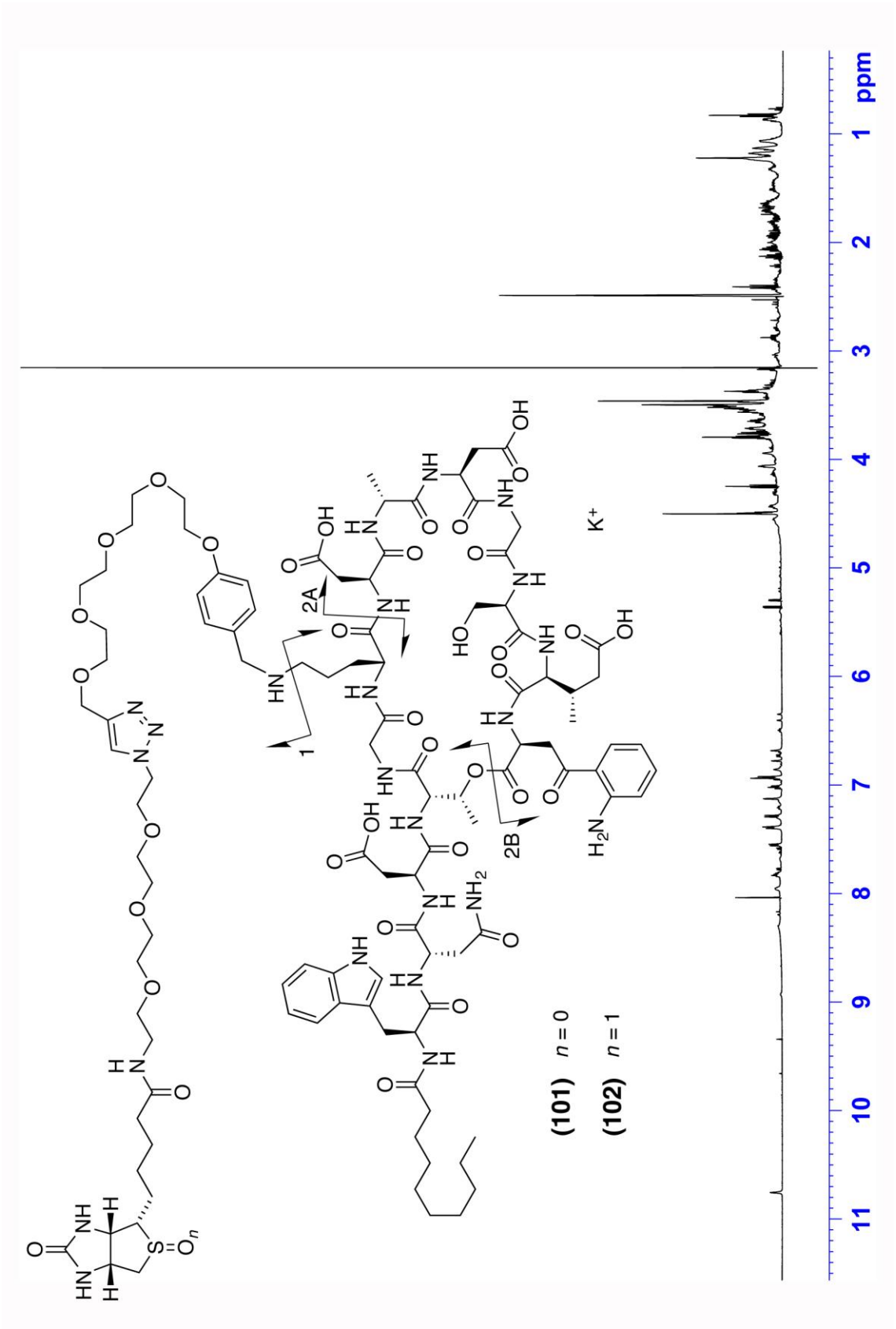


8.2.1.6 ROESY spectrum (500 ms) of biotin-manzamine (98; $CDCl_3$, 600 MHz)

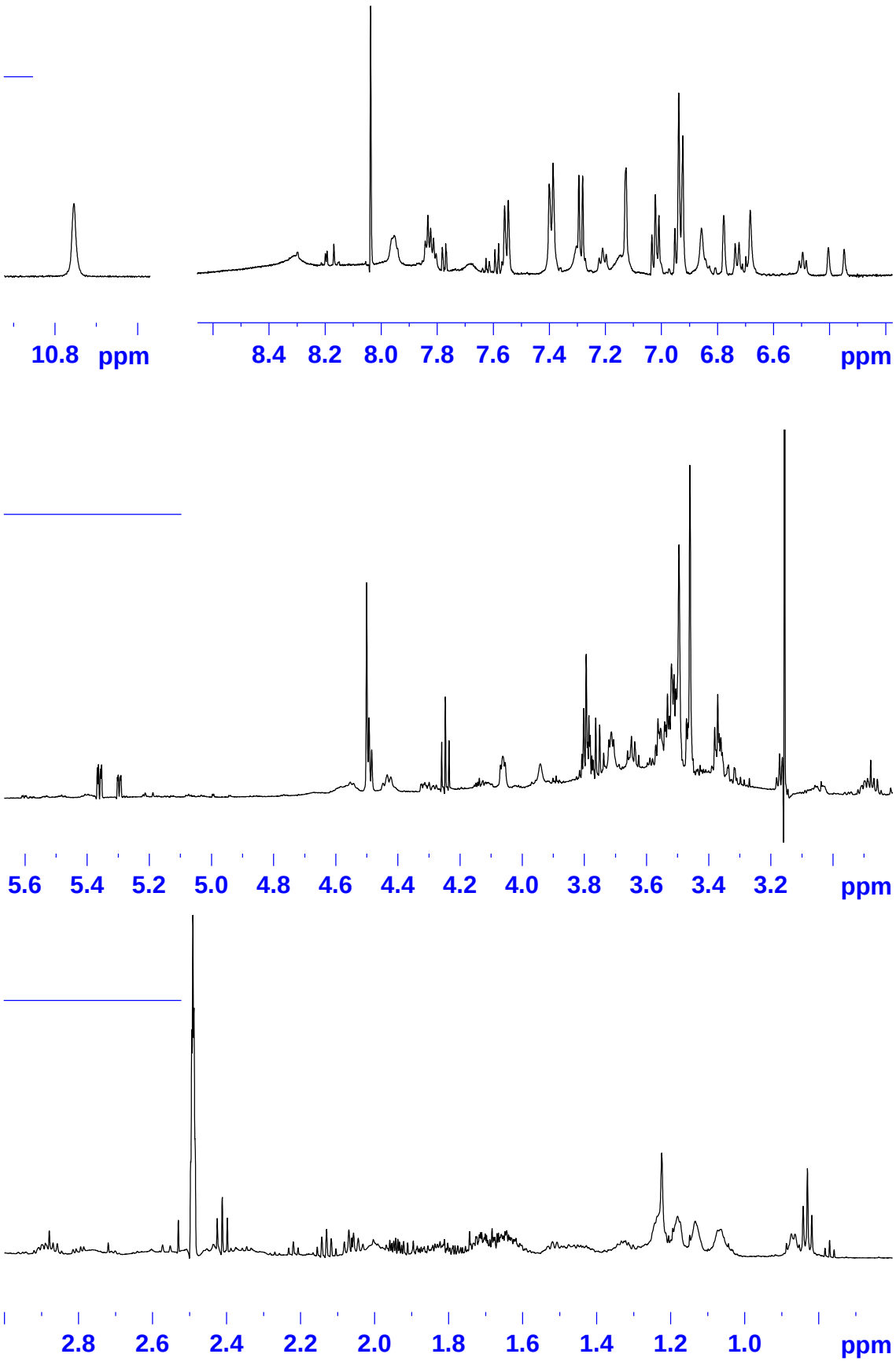


8.2.2 NMR spectra of biotin-daptomycin (101)

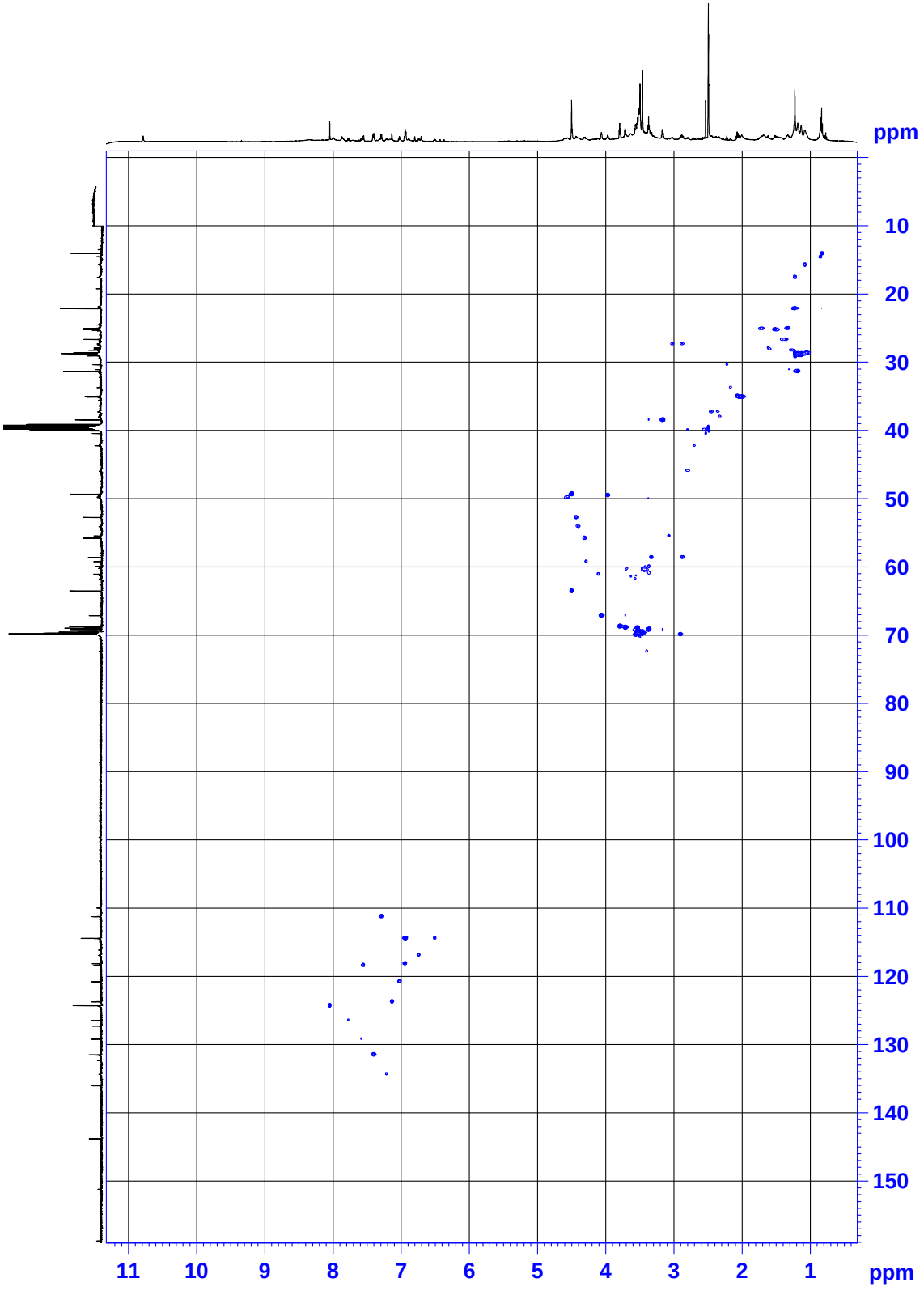
8.2.2.1 ^{1}H NMR spectrum of biotin-daptomycin (101; DMSO-d_6, 600 MHz)



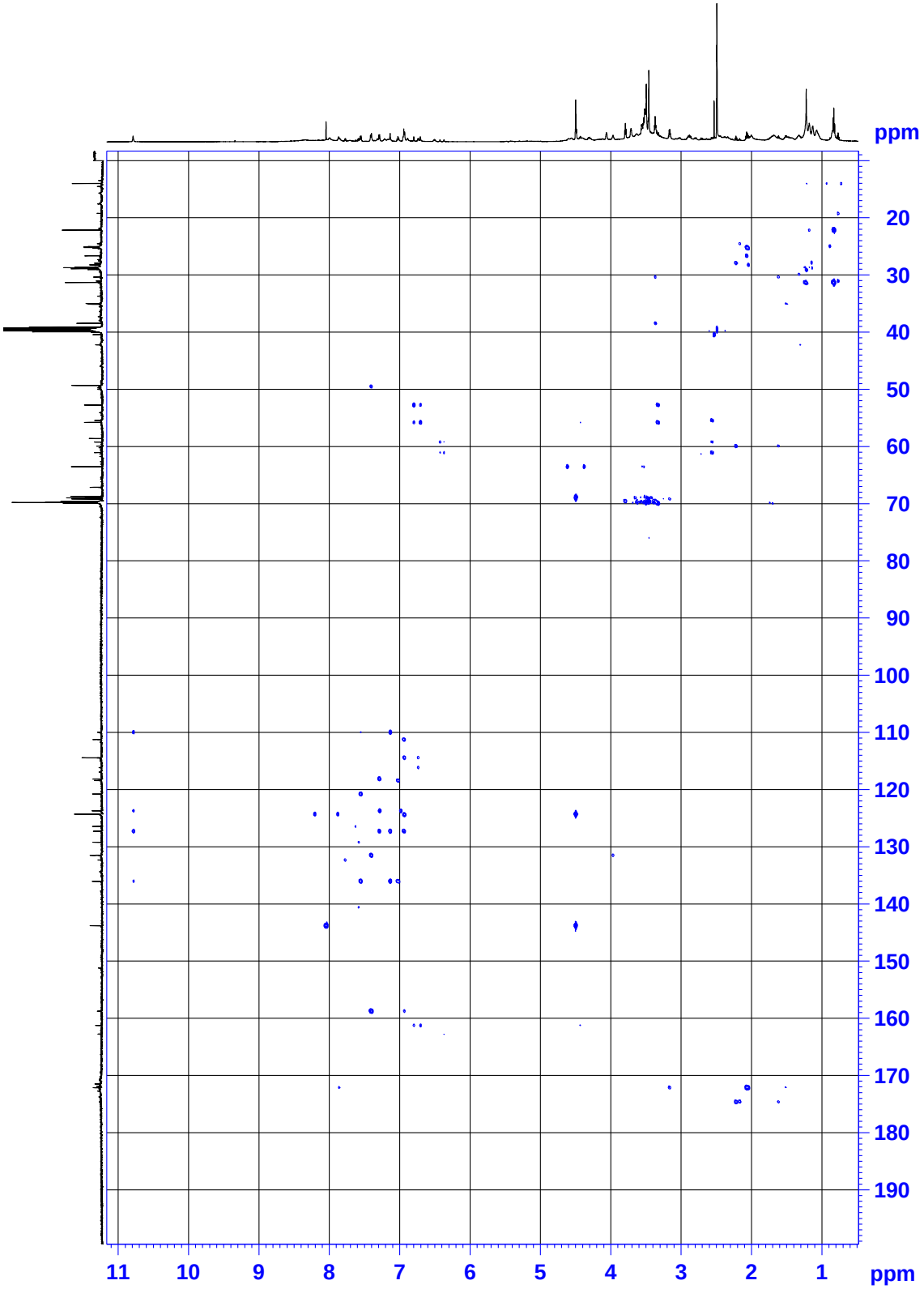
8.2.2.2 ^{1}H NMR spectrum of biotin-daptomycin (101; DMSO-d_6, 600 MHz) – expansions



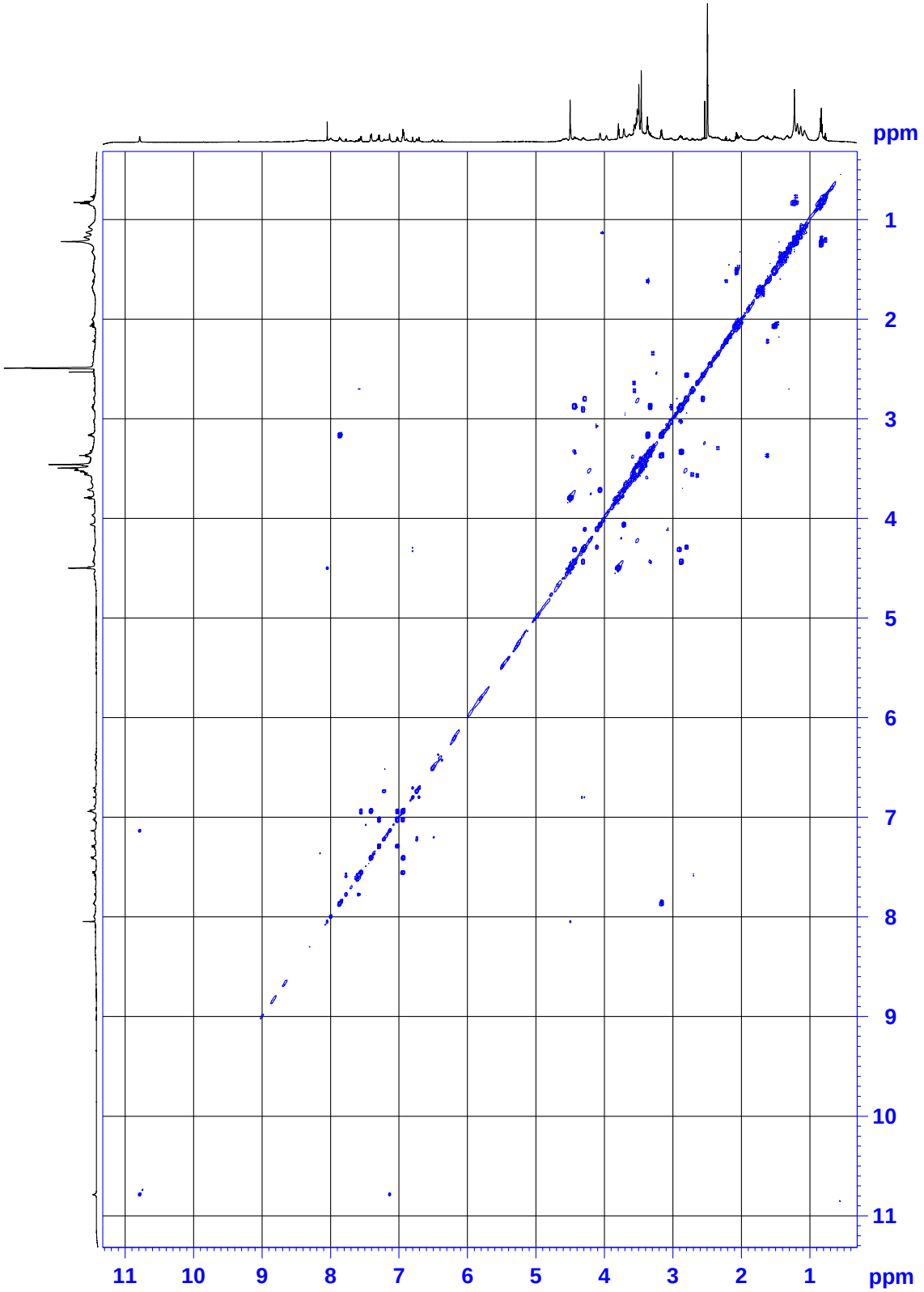
8.2.2.3 HSQC spectrum of biotin-daptomycin (101; DMSO-d_6, 600 MHz)



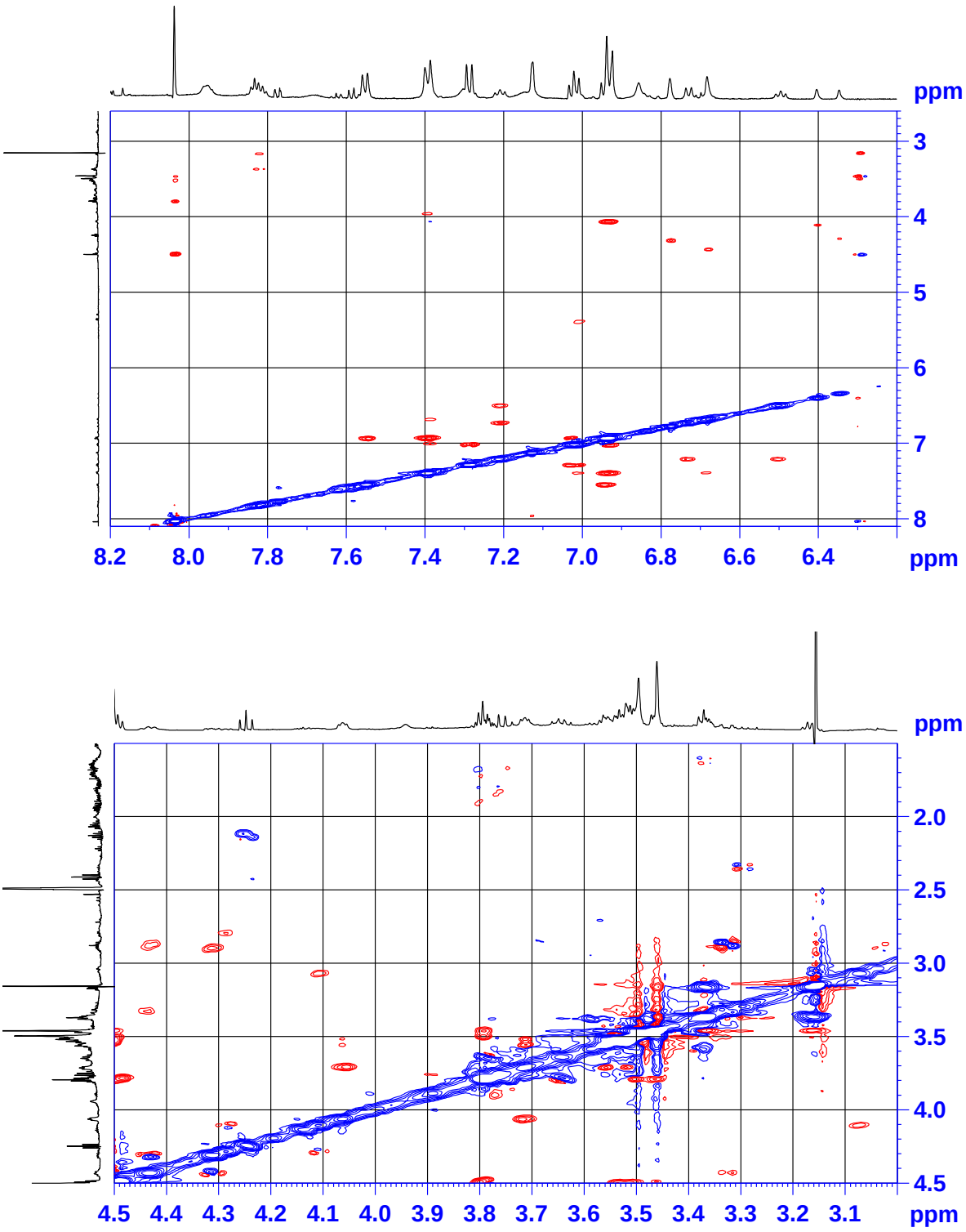
8.2.2.4 HMBC spectrum of biotin-daptomycin (101; DMSO-d_6, 600 MHz)



8.2.2.5 COSY spectrum of biotin-daptomycin (101; DMSO-d_6, 600 MHz)

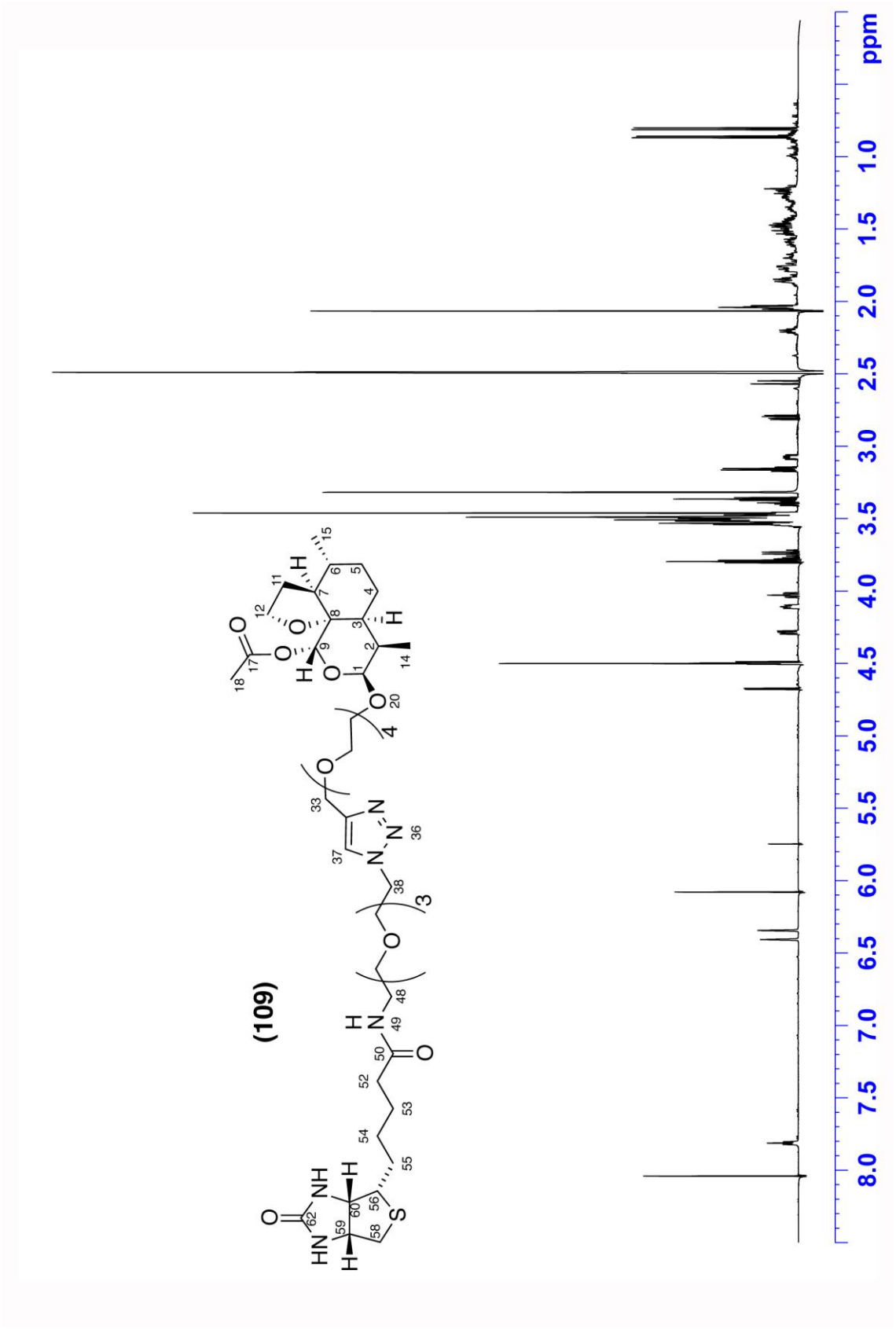


8.2.2.6 ROESY spectrum (250 ms) of biotin-daptomycin (101; DMSO-d_6, 600 MHz)

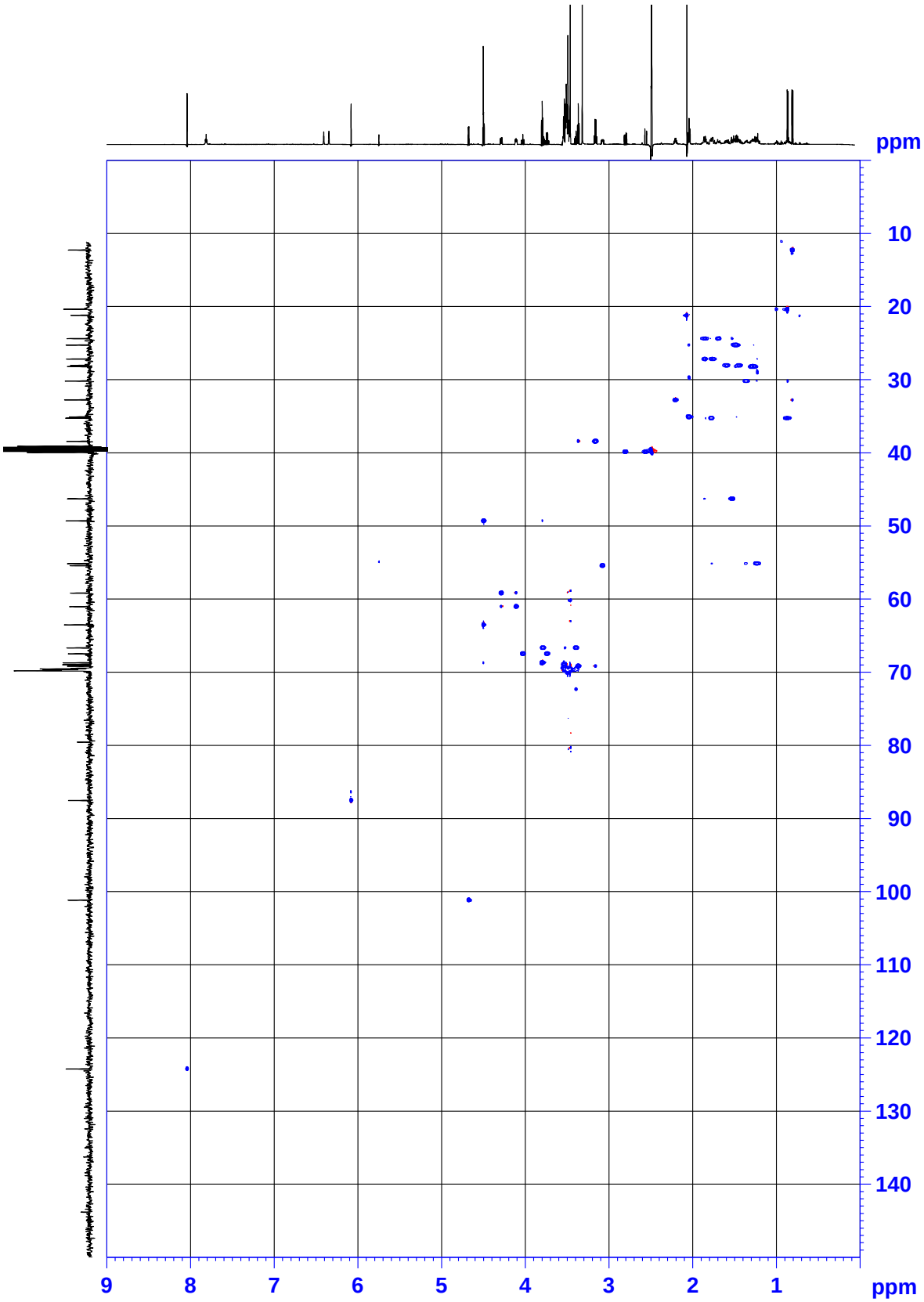


8.2.3 NMR spectra of biotin-artemisinin derivative (109)

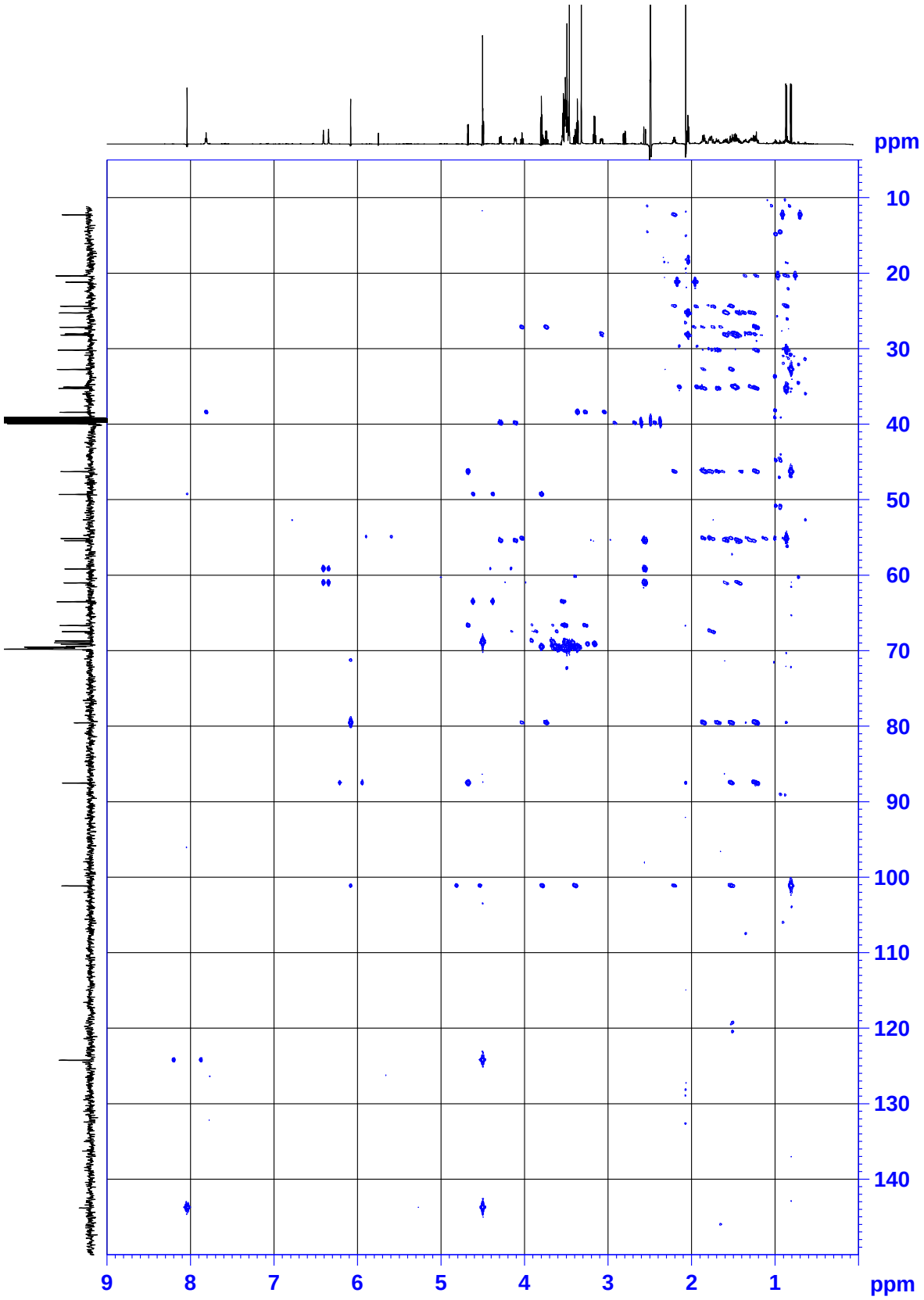
8.2.3.1 ^{1}H NMR spectrum of 109 ($CDCl_3$, 600 MHz)



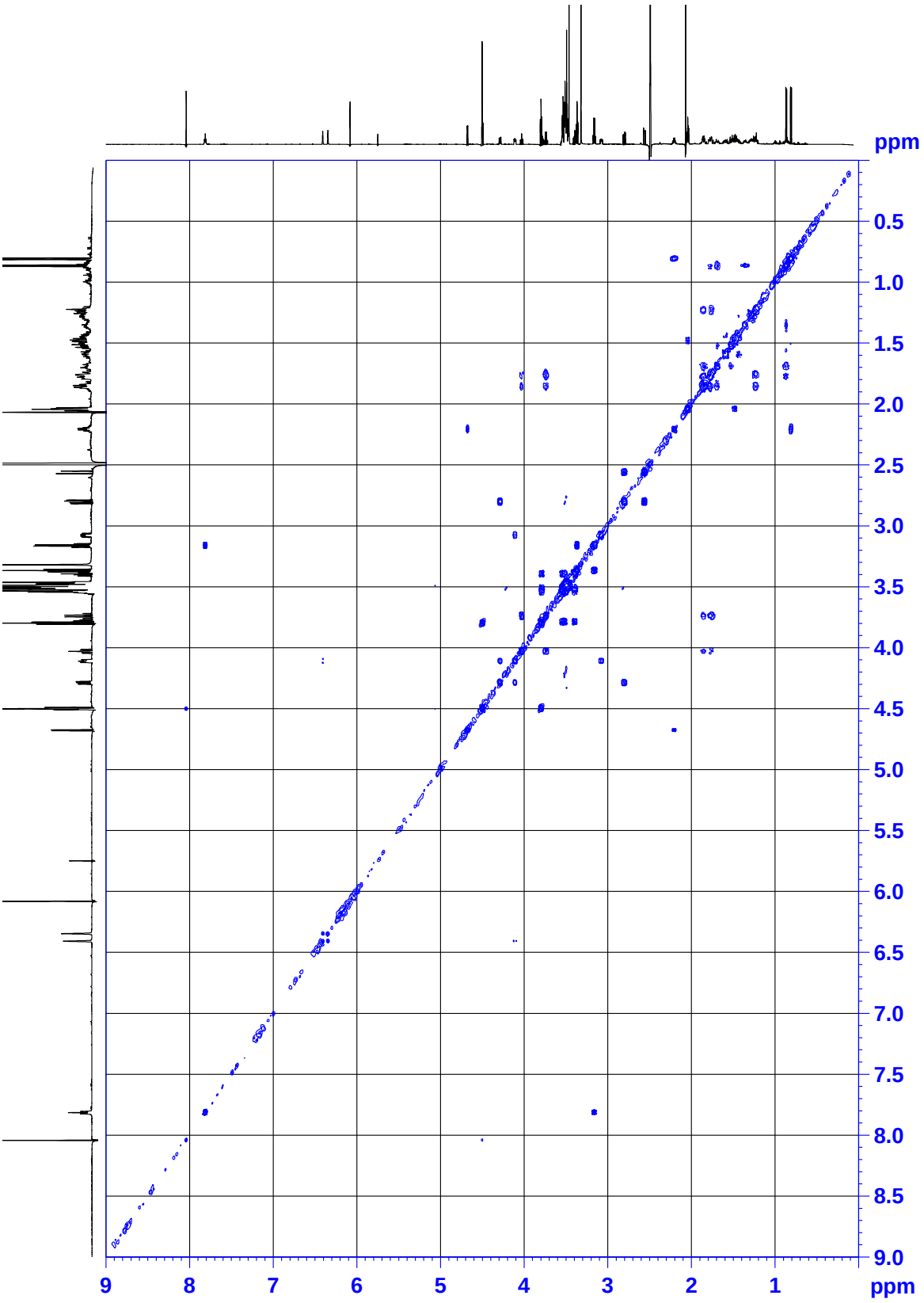
8.2.3.2 HSQC spectrum of 109 ($CDCl_3$, 600 MHz)



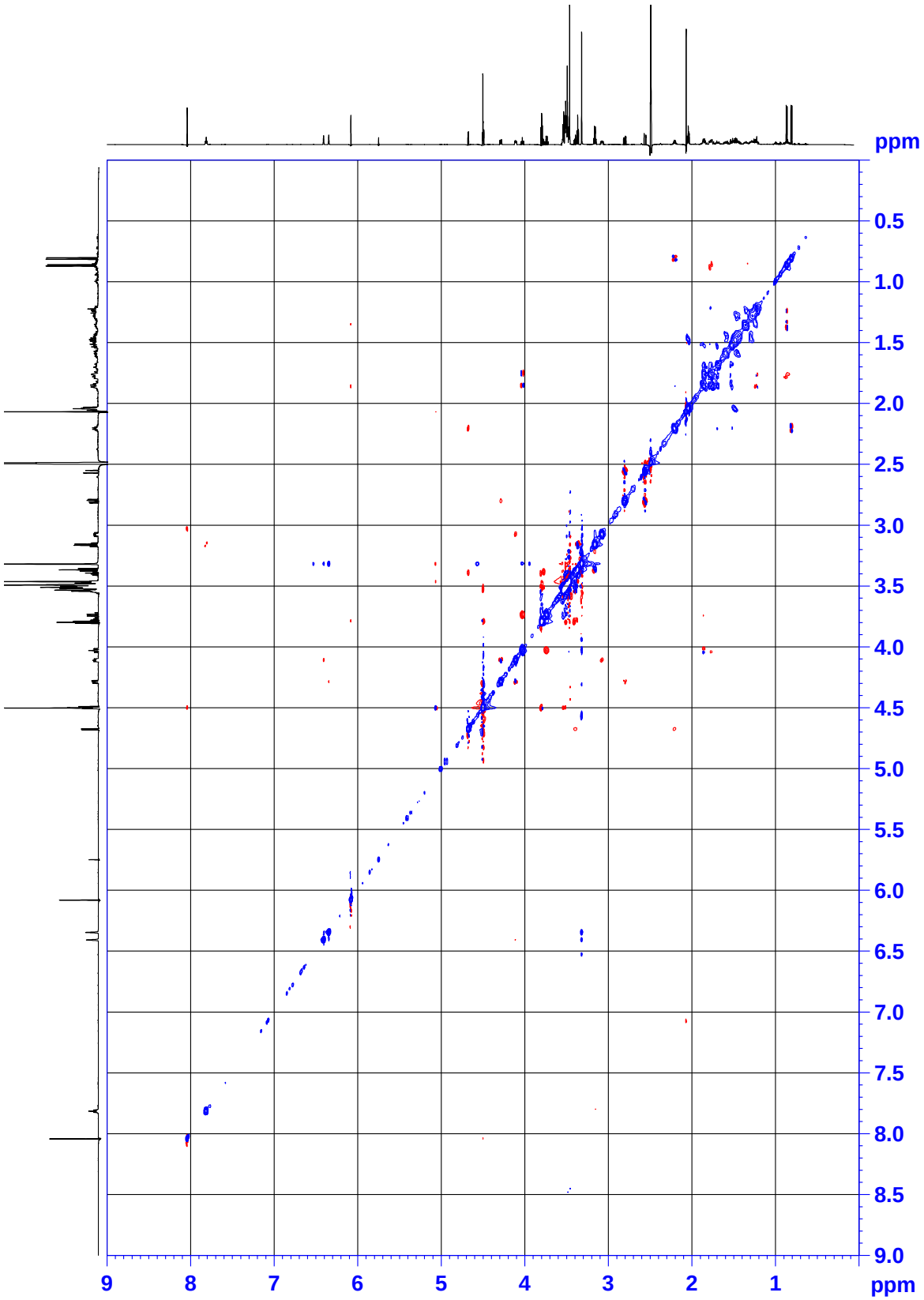
8.2.3.3 HMBC spectrum of 109 ($CDCl_3$, 600 MHz)



8.2.3.4 COSY spectrum of 109 ($CDCl_3$, 600 MHz)

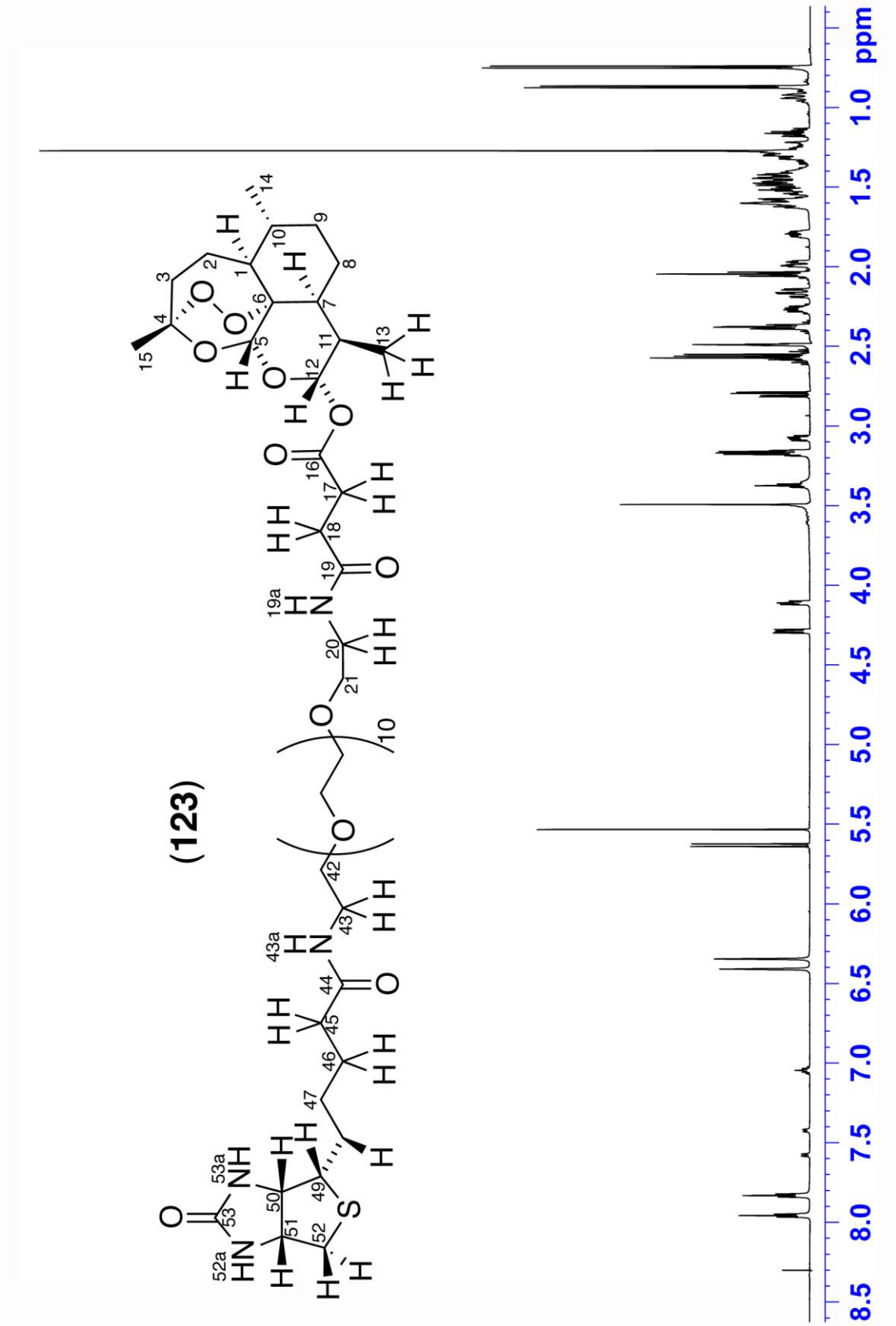


8.2.3.5 ROESY spectrum of 109 ($CDCl_3$, 600 MHz)

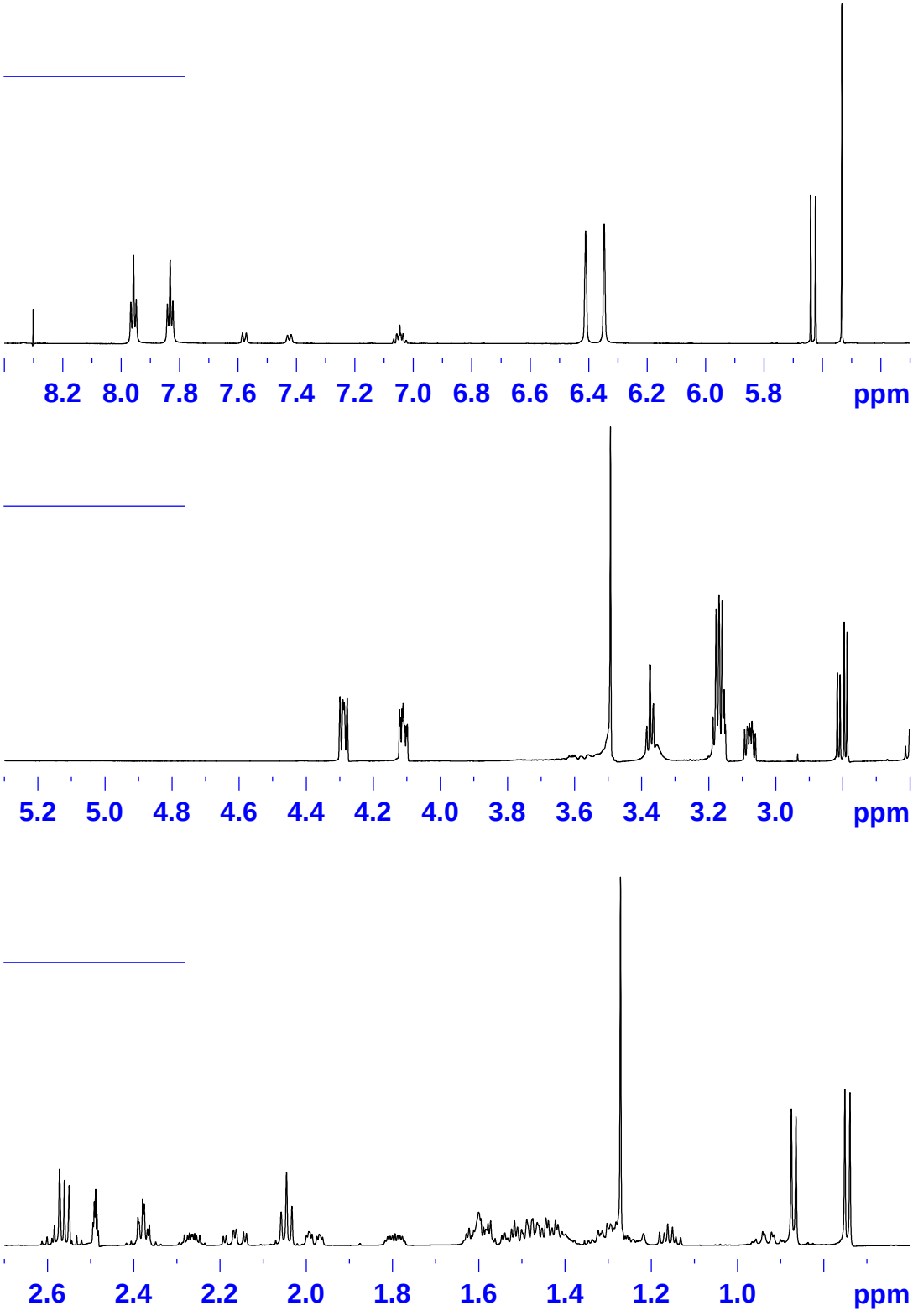


8.2.4 NMR spectra of biotin-artesunate (123)

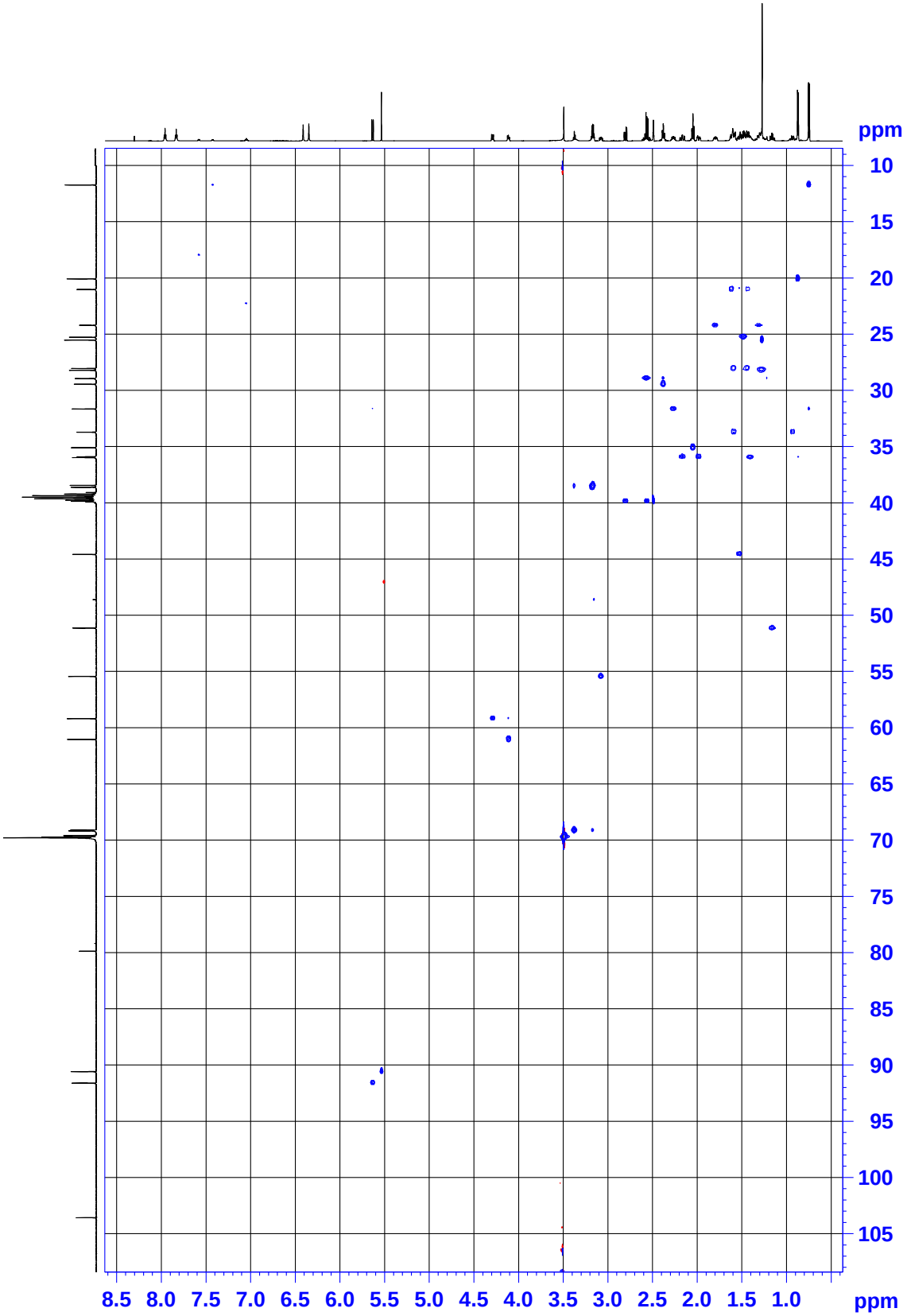
8.2.4.1 ^{1}H NMR spectrum of biotin-artesunate (123; DMSO-d_6, 600 MHz)



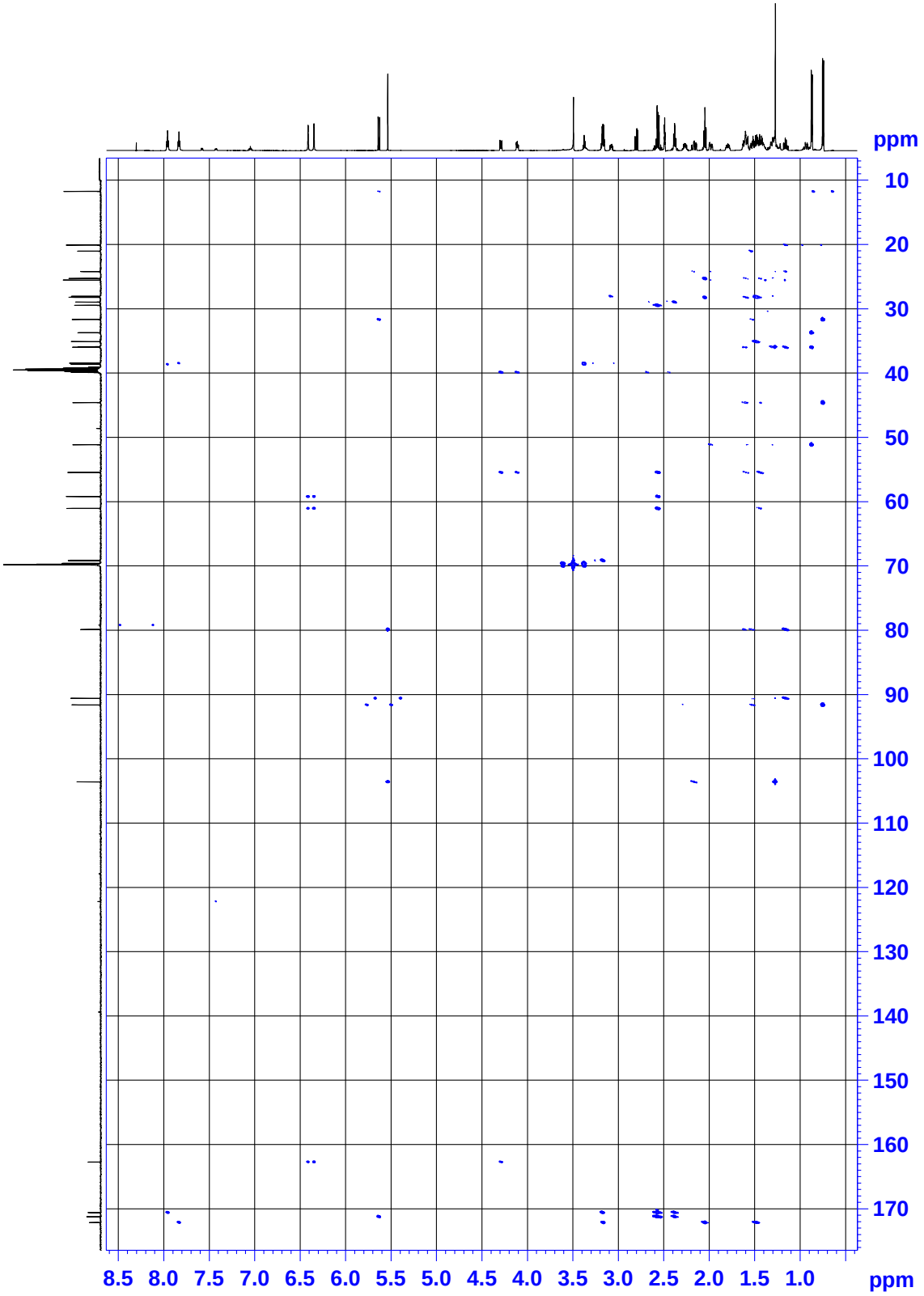
8.2.4.2 ^{1}H NMR spectrum of biotin-artesunate (123; DMSO-d_6, 600 MHz) – expansions



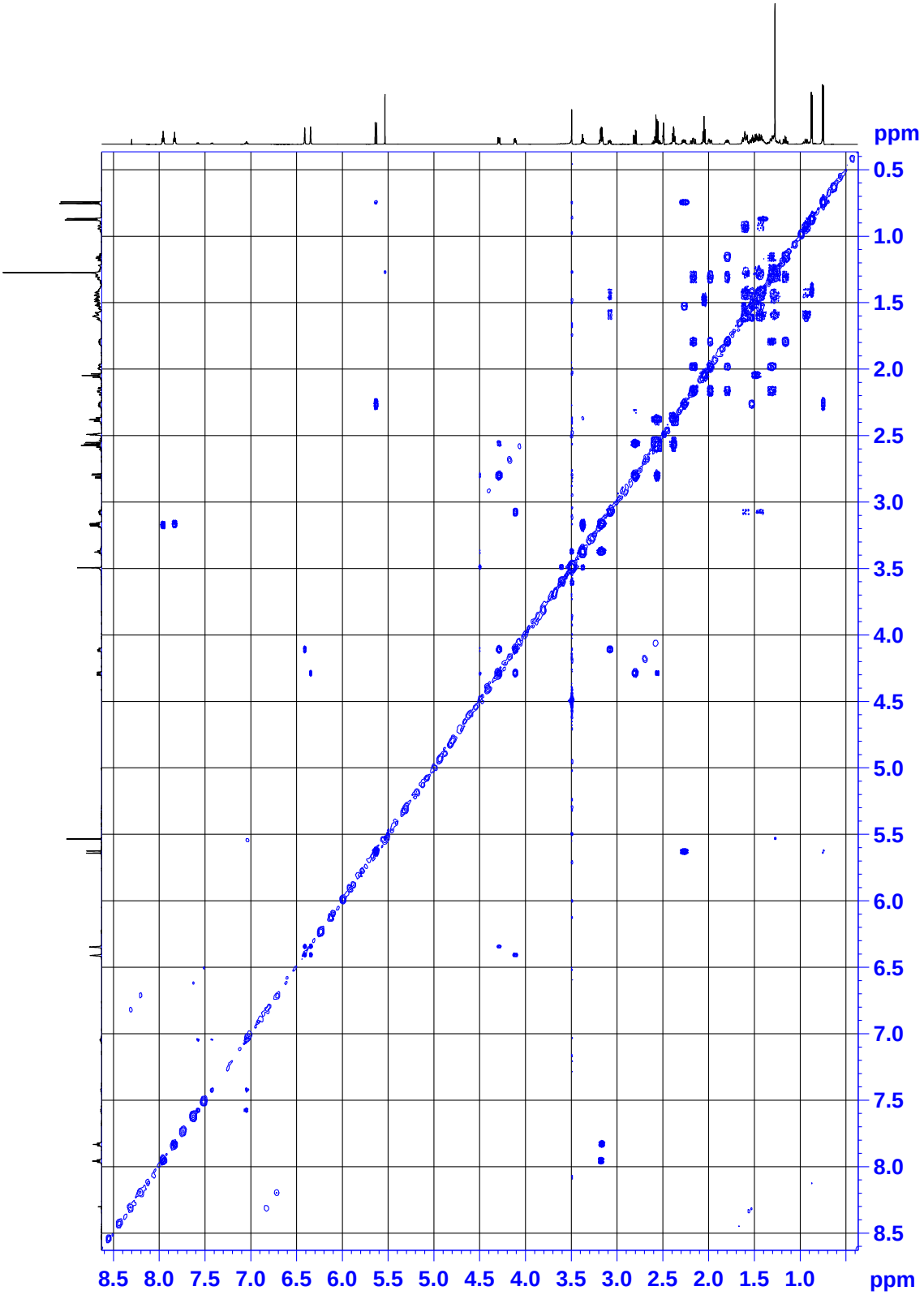
8.2.4.3 HSQC spectrum of biotin-artesunate (123; DMSO-d_6, 600 MHz)



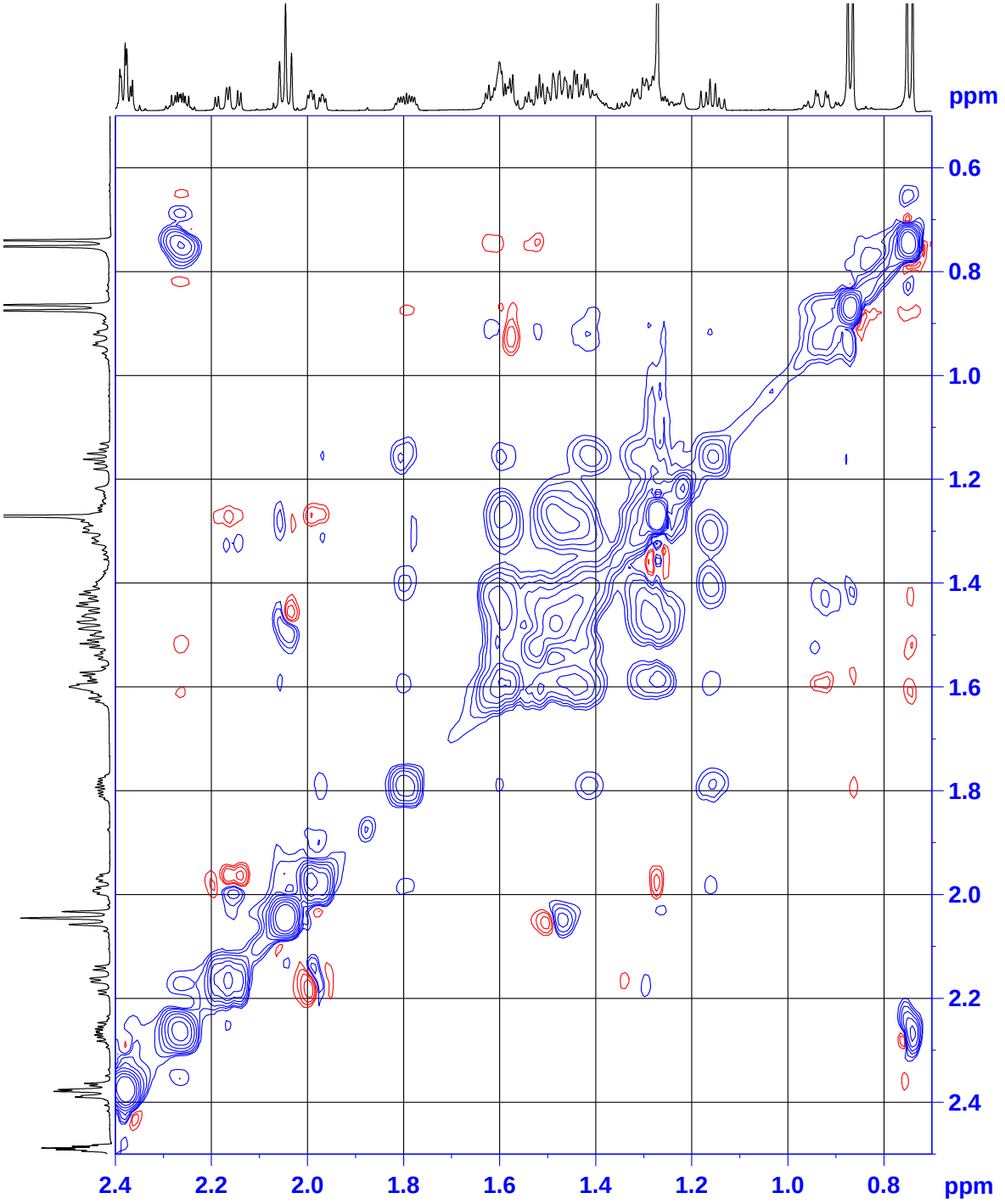
8.2.4.4 HMBC spectrum of biotin-artesunate (123; DMSO-d_6, 600 MHz)

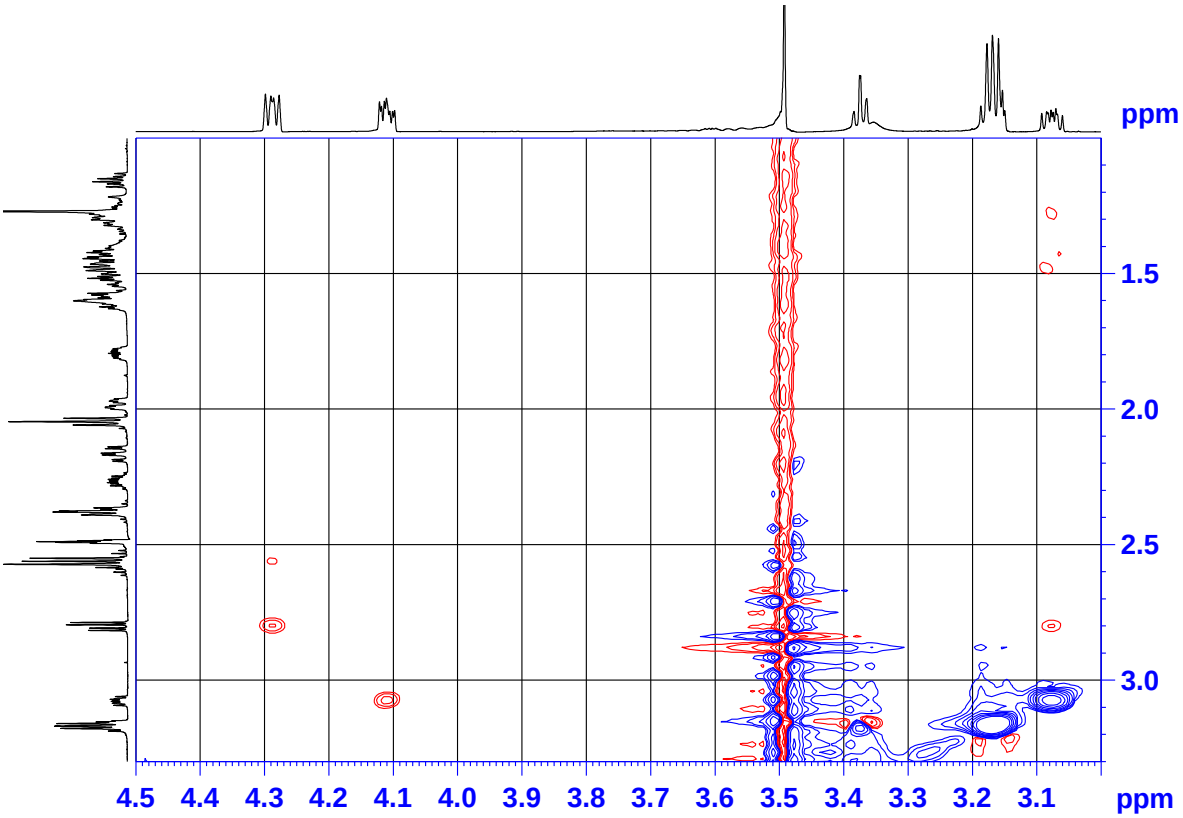


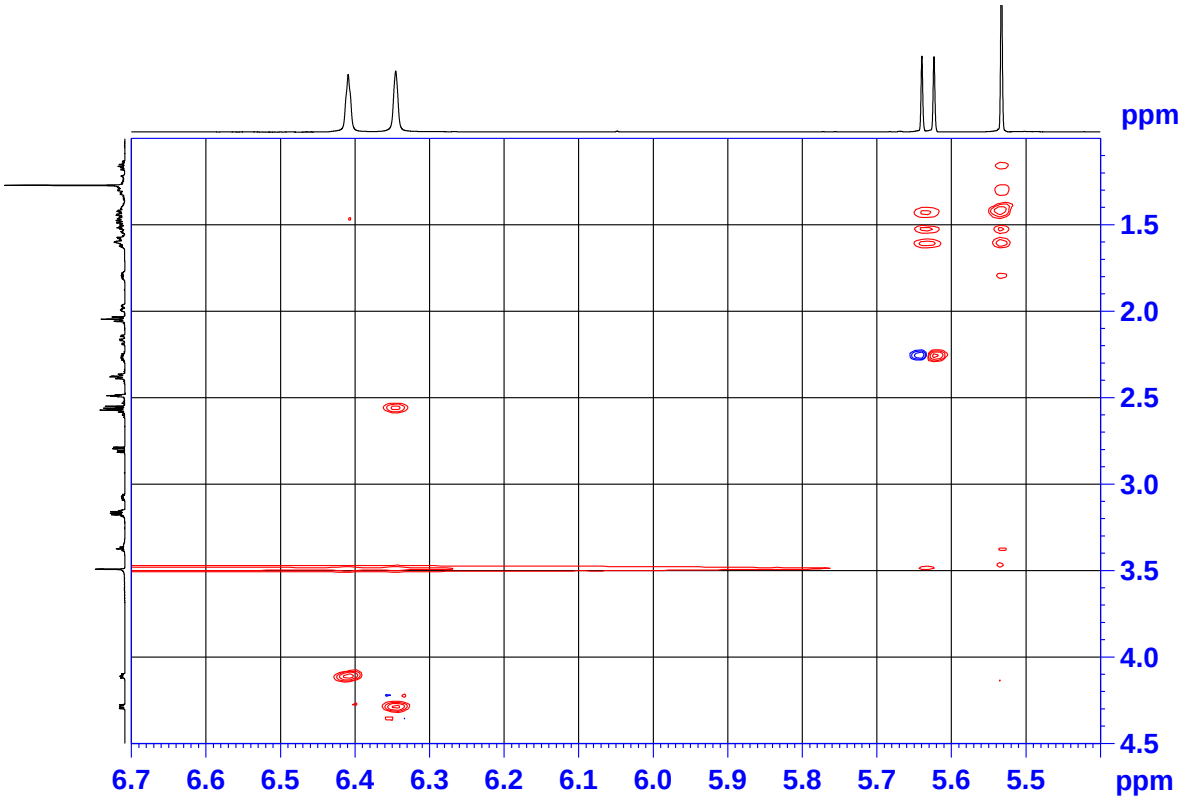
8.2.4.5 COSY spectrum of biotin-artesunate (123; DMSO-d_6, 600 MHz)



8.2.4.6 ROESY spectrum of biotin-artesunate (123; DMSO-d_6, 600 MHz)




ppm
1.5
2.0
2.5
3.0
4.5 4.4 4.3 4.2 4.1 4.0 3.9 3.8 3.7 3.6 3.5 3.4 3.3 3.2 3.1
ppm


ppm
1.5
2.0
2.5
3.0
3.5
4.0
4.5
6.7 6.6 6.5 6.4 6.3 6.2 6.1 6.0 5.9 5.8 5.7 5.6 5.5
ppm

8.3 Display cloning

8.3.1 Physical data for PCR primers

PCR primers were designed by Dr. Andrew Piggott with the Primer3 software package[399] using the following criteria:

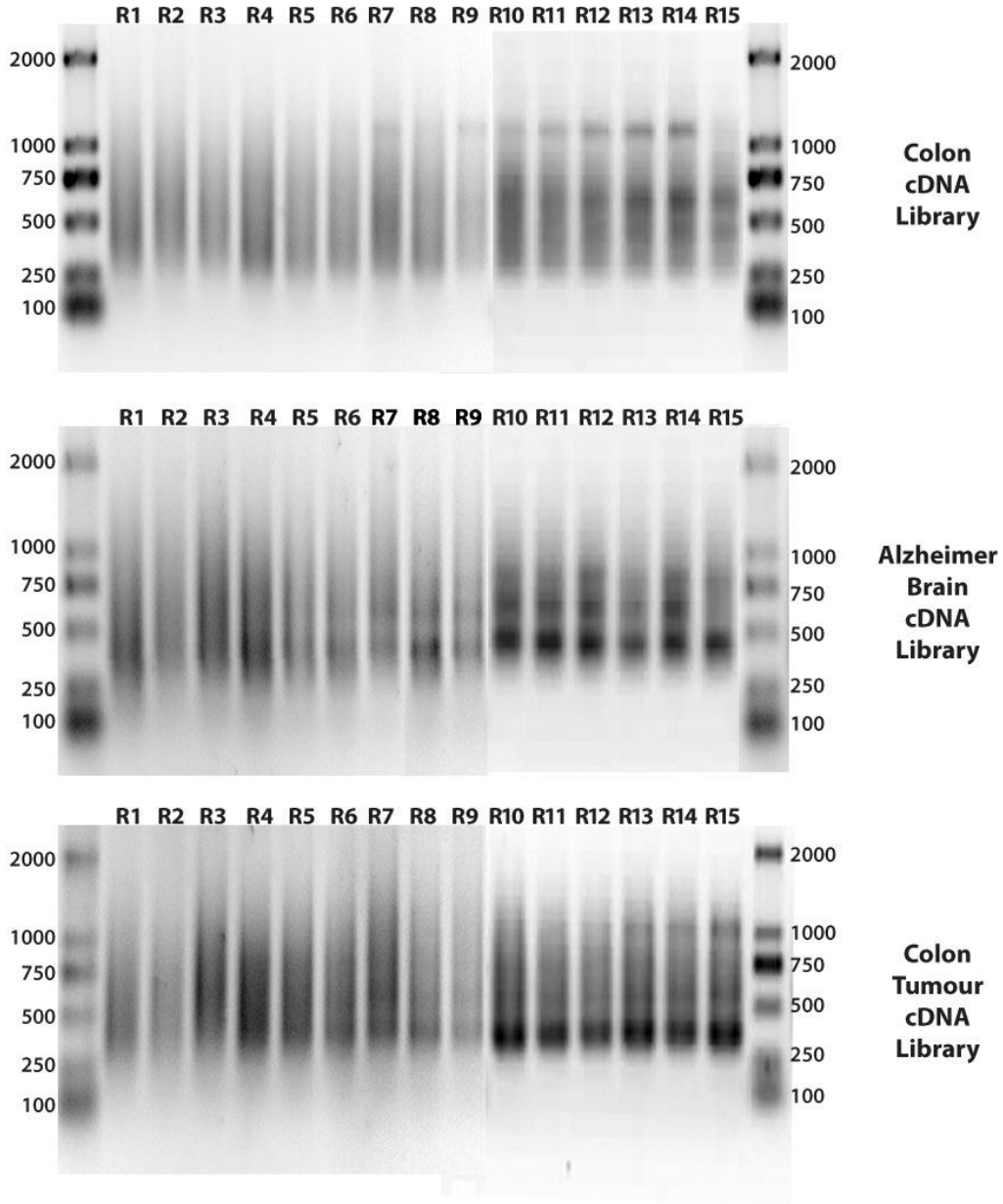
- Optimum length = 20 nucleotides
- Optimum melting temperature = 60 °C
- Optimum %GC = 50%
- Maximum base mismatch = 0

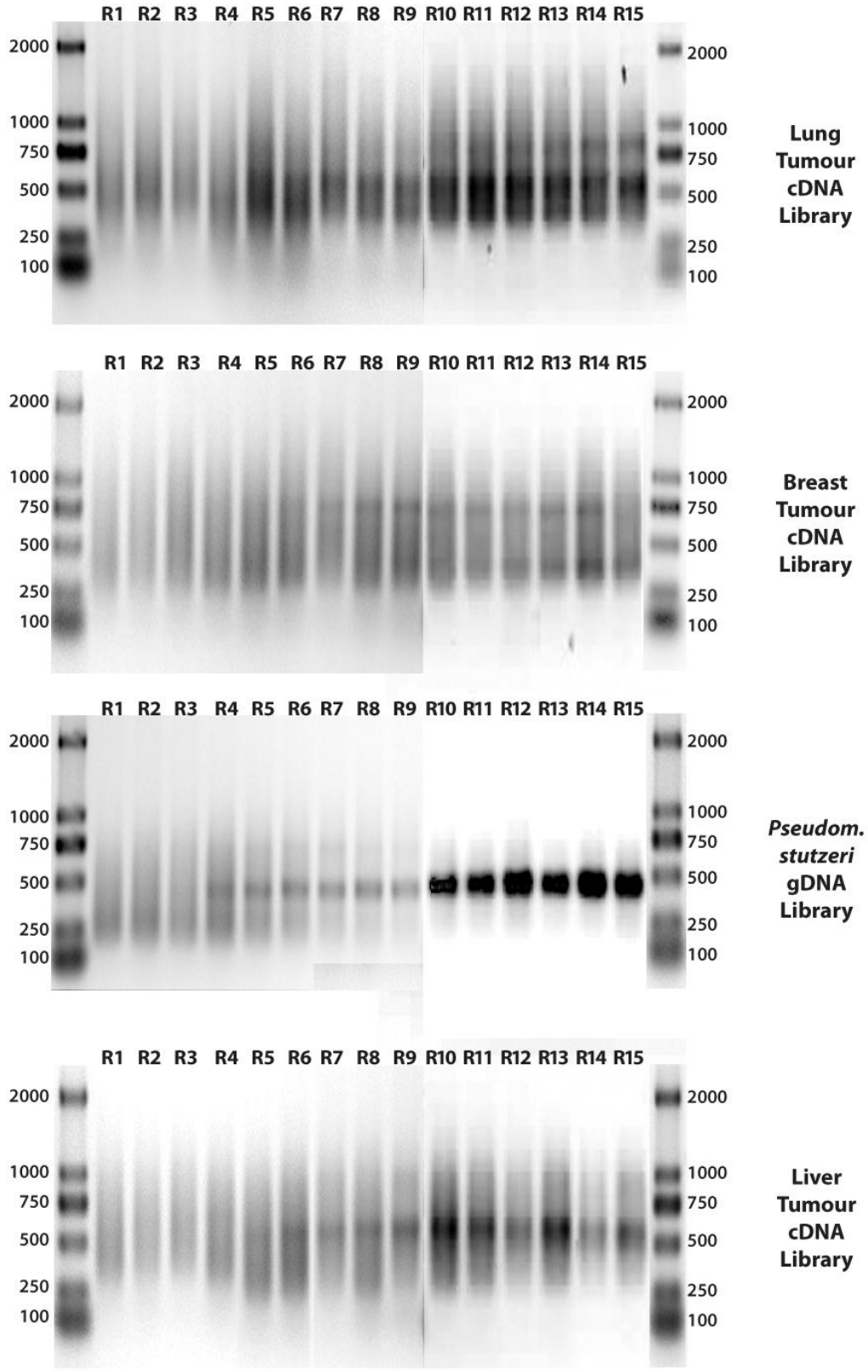
Primer	Sequence (5'3')	MW (Da)	Tm (°C)	% GC	Product Size (bp)	
T7	UP	GGAGCTGTCGTATTCCAGTC	6124	61	55	N/A
	DOWN	AACCCCTCAAGACCCGTTTA	6006	64	50	
FKBP 1a	FWD	GGGATGCTTGAAGATGGAAA	6270	64	45	216
	REV	GAAGACGAGAGTGGCATGTG	6271	64	55	
FKBP 1b	FWD	ACAGGAATGCTCCAAAATGG	6159	64	45	231
	REV	CAGCTCCACGTCAAAGATGA	6095	64	50	
FKBP 2	FWD	GAAGCTGGAAGATGGGACAG	6280	64	55	234
	REV	ATTTTGAGCAGCTCCACCTC	6028	63	50	
FKBP 3	FWD	GGCCAAGACAGCTAACAAGG	6169	64	55	475
	REV	AATTTTGGCATCAGGCTGTC	6123	64	45	

8.3.2 Agarose gels from round-15 biopanning

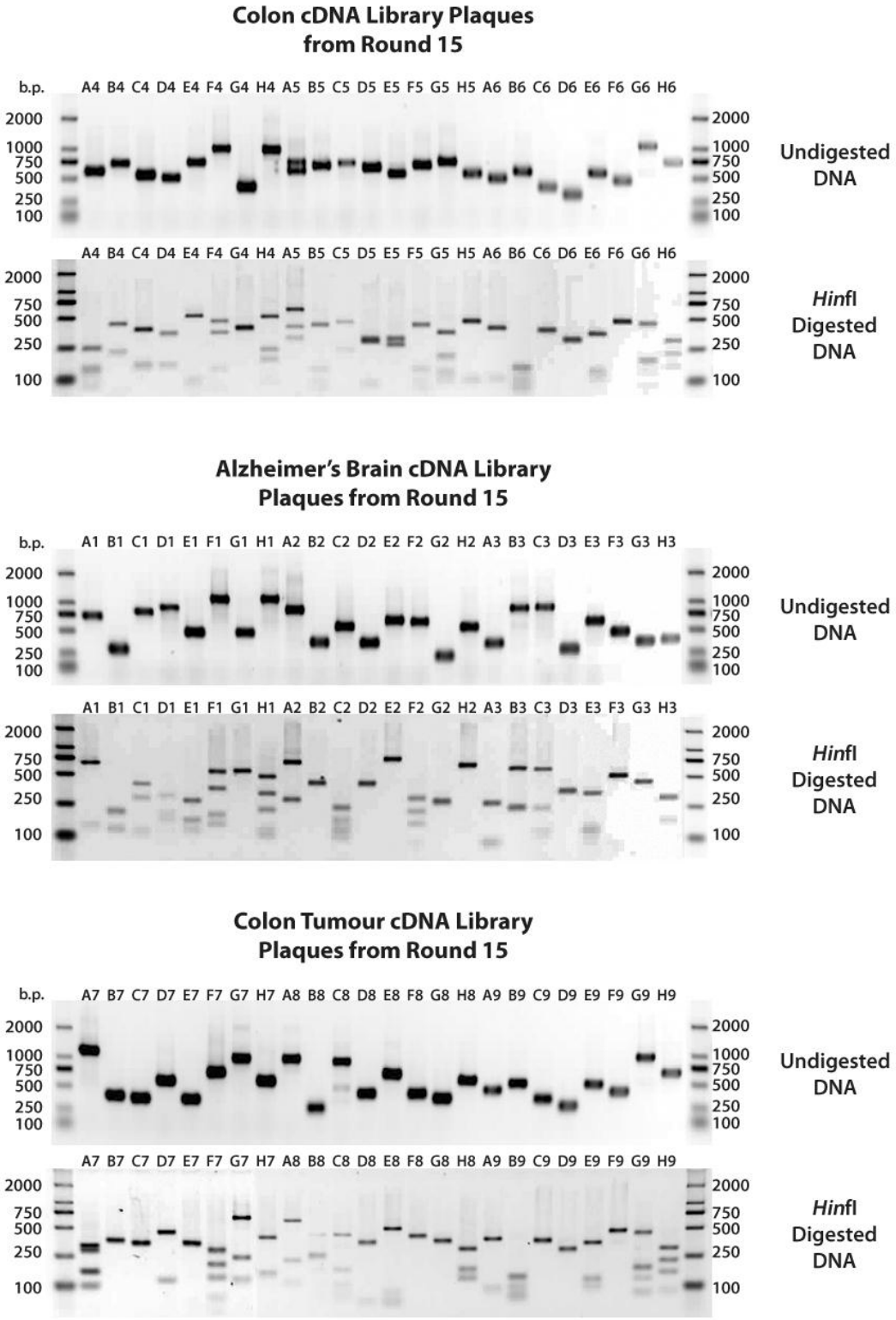
8.3.2.1 Biotin-manzamine

8.3.2.1.1 Agarose gels of PCR products of sublibraries 1-15

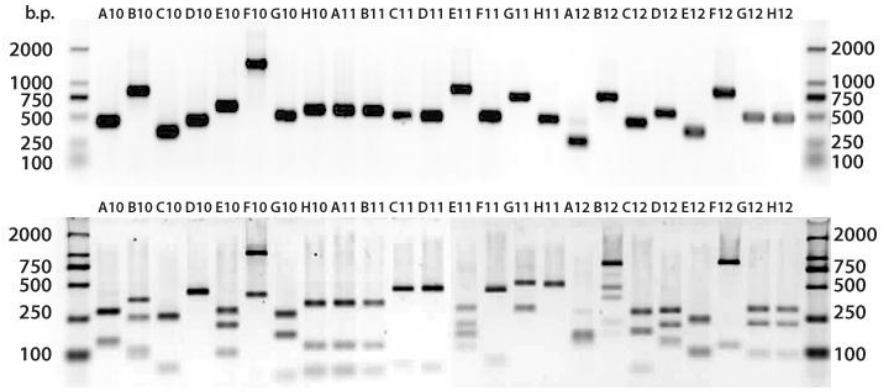



R1 R2 R3 R4 R5 R6 R7 R8 R9 R10 R11 R12 R13 R14 R15
2000
1000
750
500
250
100
Lung Tumour cDNA Library
Breast Tumour cDNA Library
Pseudom. stutzeri gDNA Library
Liver Tumour cDNA Library

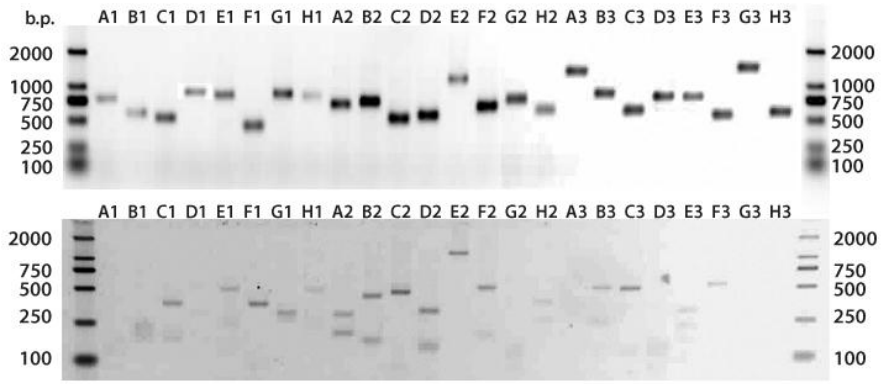
8.3.2.1.2 Agarose gels of single plaques picked from round-15 sublibraries



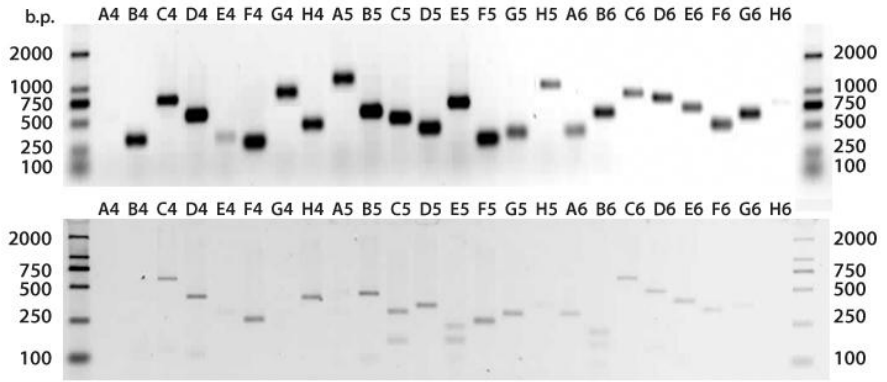
Liver Tumour cDNA Library
Plaques from Round 15

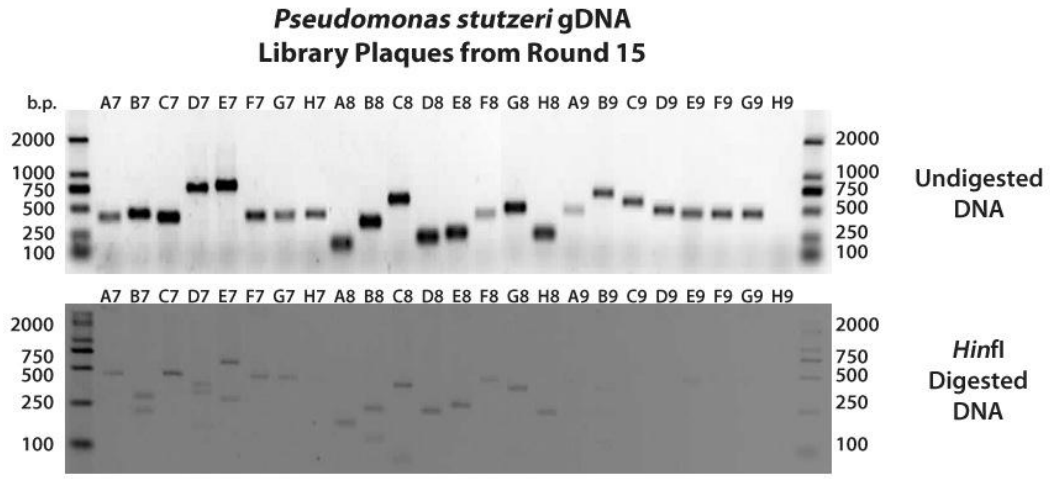


Lung Tumour cDNA Library
Plaques from Round 15



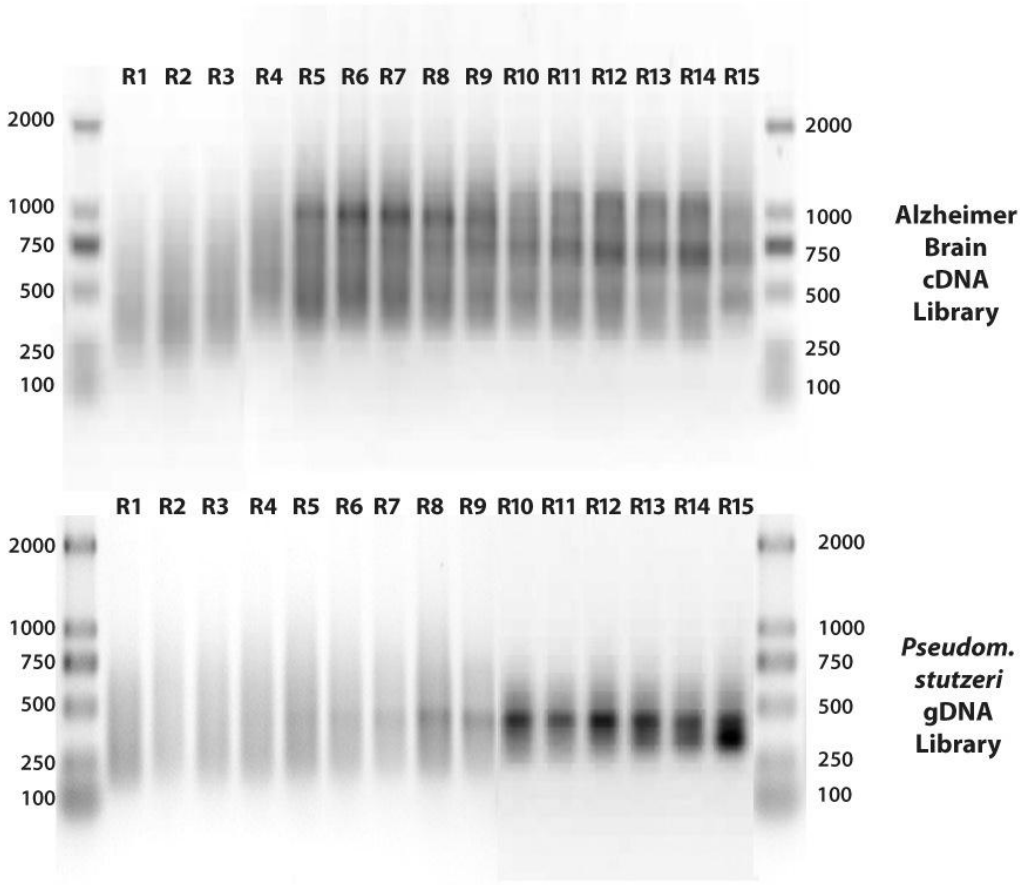
Breast Tumour cDNA Library
Plaques from Round 15



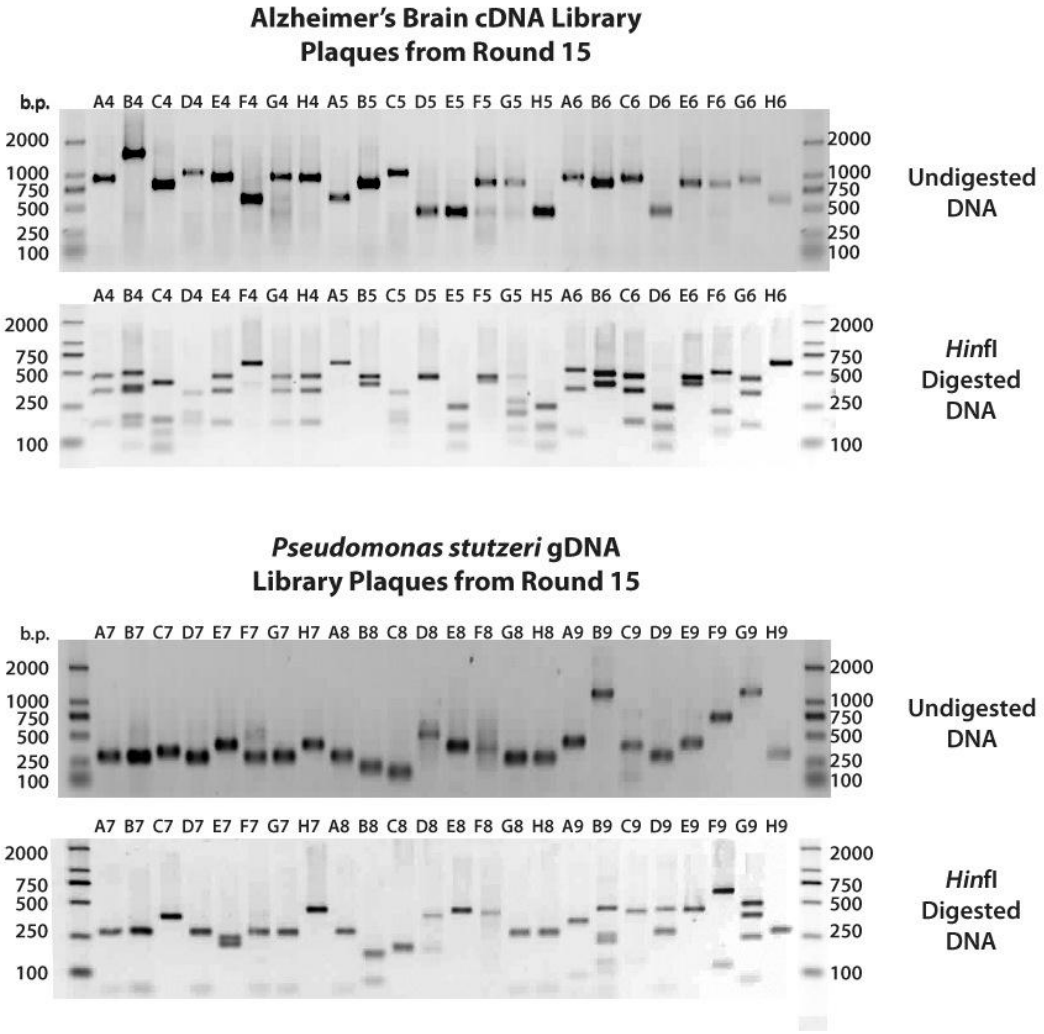


8.3.2.2 Biotin-daptomycin

8.3.2.2.1 Agarose gels of PCR products of sublibraries 1-15



8.3.2.2.2 Agarose gels of single plaques picked from round-15 sublibraries



8.3.3 DNA sequences of rescued clones

If not otherwise stated, all the cDNA inserts of all rescued clones were analysed by a BLASTn search (at http://blast.ncbi.nlm.nih.gov/) with the following parameters: Database - Reference RNA sequences (refseq_rna); Organism - Homo sapiens (taxid:9606); Program Selection - Highly similar sequences (megablast).

8.3.3.1 Biotinylated FK506

8.3.3.1.1 Human brain

Brain cDNA Library – Round 7 – Plaque B1

TTTCTTTGGGGAACG**GAATTCA**AGCGGTCGCTGTTGGTCCACGCCGCCCGTCGCGCCGCCCGCCCGCTCAGCG
TCCGCCGCCGCCATGGGAGTGCAGGTGGAAACCATCTCCCCAGGAGACGGGCGCACCTTCCCCAAGCGCGGCC
AGACCTGCGTGGTGCACTACACCGGGATGCTTGAAGATGGAAAGAAATTTGATTCCTCCCGGGACAGAAACAA
GCCCTTTAAGTTTATGCTAGGCAAGCAGGAGGTGATCCGAGGCTGGGAAGAAGGGGTTGCCCAGATGAGTGTG
GGTCAGAGAGCCAAACTGACTATATCTCCAGATTATGCCTATGGTGCCACTGGGCACCCAGGCATCATCCCAC
CACATGCCACTCTCGTCTTCGATGTGGAGCTTCTAAAACTGGAATGACAGGAATGGCCTCCTCCCTTAGCTCC
CGAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTG (up primer)

LOCUS NM_054014 901 bp mRNA linear PRI 08-OCT-2011
DEFINITION **Homo sapiens FK506 binding protein 1A, 12kDa (FKBP1A), transcript variant 2, mRNA.**
ACCESSION NM_054014 VERSION NM_054014.3 GI:315139018

CDS 175-501
```
  1 tttccgggac gtcgcgcgcc gtgtggggcg cgcacgcagg gctgggcgtg agggggcgtg
 61 cgcgtgcgca ggcgacgcgc cgaggtacta ggcagagccg tggaaccgcc gccaggtcgc
121 tgttggtcca cgccgcccgt cgcgccgccc gcccgctcag cgtccgccgc cgcc atggga
181 gtgcaggtgg aaaccatctc cccaggagac gggcgcacct tccccaagcg cggccagacc
241 tgcgtggtgc actacaccgg gatgcttgaa gatggaaaga aatttgattc ctcccgggac
301 agaaacaagc cctttaagtt tatgctaggc aagcaggagg tgatccgagg ctgggaagaa
361 ggggttgccc agatgagtgt gggtcagaga gccaaactga ctatatctcc agattatgcc
421 tatggtgcca ctgggcaccc aggcatcatc ccaccacatg ccactctcgt cttcgatgtg
481 gagcttctaa aactggaatg acaggaatgg cctcctccct tagctccctg ttcttgggta
541 aggaaatgga atactgaagg gcccttcact gcctttgctc ctcccatgtt atgcccagcg
601 tttgatgggt agcagagaga acaaaaaaca ccacaaggct atttttcccc ctgcattctt
661 tctgtattga gtatcctttc agtgttatta gtgtatgctt tgaatgtaaa aattggtcac
721 cctaaggaaa ggaattggca tgtgtatgtt cccagttcaa ctcatggaga tggcagctgt
781 ttaaatgttt ttctatgtag tttataaatt aaaactgaat tgaggactat ggaaatgtag
841 gccaaatttg tagtgccaac attttagttc tttggaaata agactcttaa tgaatgactt
901 t //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SGRCWSTPPVAPPARSASAAAMGVQVETISPGDGRTFPKRGQTCVVHYTGMLED GKKFDSSRDRNKPFKFMLGKQEVIRGWEEGVAQMSVGQRAKLTISPDYAYGATG HPGIIPPHATLVFDVELLKLE**
Frame 2: AVAVGPRRPSRRPPAQRPPPPWECRWKPSPQETGAPSPSAARPAWCTTPGCLKM ERNLIPPGTETSPLSLC
Frame 3: RSLLVHAARRAARPLSVRRRHGSAGGNHLPRRRAHLPQARPDLRGALHRDA

Brain cDNA Library – Round 7 – Plaque E1

ATTAAGGATAAAC**GAATTCA**AGCCGGAGCCGAGCCGGGGTCGGGCAGCAGCAGGGACCCCCCAGAGGCGGGGC
CTGTGGGACCGCTATGGGCGTGGAGATCGAGACCATCTCCCCCGGAGACGGAAGGACATTCCCCAAGAAGGGC
CAAACGTGTGTGGTGCACTACACAGGAATGCTCCAAAATGGGAAGAAGTTTGATTCATCCAGAGACAGAAACA
AACCTTTCAAGTTCAGAATTGGCAAACAGGAAGCCATCAAAGGTTTTGAAGAGGGTGCAGCCCAGAT (up
primer)

LOCUS NM_004116 971 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens FK506 binding protein 1B, 12.6 kDa (FKBP1B), transcript variant 1, mRNA.**
ACCESSION NM_004116 VERSION NM_004116.3 GI:206725519

CDS 137-463
```
  1 gactccagcc gcacctcctc cggctctgca gtggcggcga ggaggcgagc cggagcgacg
 61 gcggggctgg ggccggagcc gagccggggt cgggcagcag cagggacccc ccagaggcgg
121 ggcctgtggg accgct**atgg gcgtggagat cgagaccatc tcccccggag acggaaggac**
181 **attccccaag aagggccaaa cgtgtgtggt gcactacaca ggaatgctcc aaaatgggaa**
241 **gaagtttgat tcatccagag acagaaacaa acctttcaag ttcagaattg gcaaacagga**
301 **agtcatcaaa ggttttgaag agggtgcagc ccagatgagc ttggggcaga gggcgaagct**
361 gacctgcacc cctgatgtgg catatggagc cacgggccac cccggtgtca tccctcccaa
421 tgccaccctc atctttgacg tggagctgct caacttagag tgaaggcagg aaggaactca
481 aggtggctgg agatggctgc tgctcaccct cctagcctgc tctgccactg ggacggctcc
541 tgcttttggg gctcttgatc agtgtgctaa cctcactgcc tcatggcatc atccattctc
601 tctgcccaag ttgctctgta tgtgttcgtc agtgttcatg cgaattcttg cttgaggaaa
661 cttcggttgc agattgaagc atttcaggtt gtgcattttg tgtgatgcat gtagtagcct
721 ttcctgatga cagaacacag atctcttgtt cgcacaatct acactgcctt accttcactt
781 aaaccacaca cacaaggtgc tcagacatga aatgtacatg gcgtaccgta cacagaggga
841 cttgagccag ttacctttgc tgtcactttc tctcttataa attctgttag ctgctcactt
901 aaacaatgtc ctctttgaga aaatgtaaaa taaaggctct gtgcttgaca aaaaaaaaaa
961 aaaaaaaaaa a //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (Bold = correct reading frame, underlined = expressed by CDS):

Frame 1: **SRSRAGVGQQQGPPRGGACGTAMGVEIETISPGDGRTFPKKGQTCVVHYTGMLQN**
GKKFDSSRDRNKPFKFRIGKQEVIKGFEEGAAQ
Frame 2: AGAEPGSGSSRDPPEAGPVGPLWAWRSRPSPPETEGHSPRRAKRVWCTTQECSKM
GRSLIHPETETNLSSSELANRKSSKVLKRVQPI
Frame 3: PEPSRGRAAAGTPQRRGLWDRYGRGDRDHLPRRRKDIPQEGPNVCGALHRNAPKW
EEV

Brain cDNA Library – Round 7 – Plaque F1

TGGTGGGGCATCG**GAATTCA**AGCGGTCGCTGTTGGTCCACGCCGCCCGTCGCGCCGCCCGCCCGCTCAGCGTC
CGCCGCCGCCATGGGAGTGCAGGTGGAAACCATCTCCCCAGGAGACGGGCGCACCTTCCCCAAGCGCGGCCAG
ACCTGCGTGGTGCACTACACCGGGATGCTTGAAGATGGAAAGAAATTTGATTCCTCCCGGGACAGAAACAAGC
CCTTTAAGTTTATGCTAGGCAAGCAGGAGGTGATCCGAGGCTGGGAAGAAGGGGTTGCCCAGATGAGTGTGGG
TCAGAGAGCCAAACTGACTATATCTCCAGATTATGCCTATGGTGCCACTGGGCACCCAGGCATCATCCCACCA
CATGCCACTCTCGTCTTCGATGTGGAGCTTCTAAAACTGGAATGACAGGAATGGCCTCCTCCCTTAGCTCCCG
AAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTAA (up
primer)

LOCUS NM_054014 901 bp mRNA linear PRI 08-OCT-2011
DEFINITION **Homo sapiens FK506 binding protein 1A, 12kDa (FKBP1A), transcript variant 2, mRNA.**
ACCESSION NM_054014 VERSION NM_054014.3 GI:315139018

CDS 175-501

```
  1 tttccgggac gtcgcgcgcc gtgtggggcg cgcacgcagg gctgggcgtg agggggcgtg
 61 cgcgtgcgca ggcgacgcgc cgaggtacta ggcagagccg tggaaccgcc gccaggtcgc
121 tgttggtcca cgccgcccgt cgcgccgccc gcccgctcag cgtccgccgc cgccatggga
181 gtgcaggtgg aaaccatctc cccaggagac gggcgcacct tccccaagcg cggccagacc
241 tgcgtggtgc actacaccgg gatgcttgaa gatggaaaga aatttgattc ctcccgggac
301 agaaacaagc cctttaagtt tatgctaggc aagcaggagg tgatccgagg ctgggaagaa
361 ggggttgccc agatgagtgt gggtcagaga gccaaactga ctatatctcc agattatgcc
421 tatggtgcca ctgggcaccc aggcatcatc ccaccacatg ccactctcgt cttcgatgtg
481 gagcttctaa aactggaatg acaggaatgg cctcctccct tagctccctg ttcttgggta
541 aggaaatgga atactgaagg gcccttcact gcctttgctc ctcccatgtt atgcccagcg
601 tttgatgggt agcagagaga acaaaaaaca ccacaaggct atttttcccc ctgcattctt
661 tctgtattga gtatcctttc agtgttatta gtgtatgctt tgaatgtaaa aattggtcac
721 cctaaggaaa ggaattggca tgtgtatgtt cccagttcaa ctcatggaga tggcagctgt
781 ttaaatgttt ttctatgtag tttataaatt aaaactgaat tgaggactat ggaaatgtag
841 gccaaatttg tagtgccaac attttagttc tttggaaata agactcttaa tgaatgactt
901 t  //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SGRCWSTPPVAPPARSASAAAMGVQVETISPGDGRTFPKRGQTCVVHYTGMLED GKKFDSSRDRNKPFKFMLGKQEVIRGWEEGVAQMSVGQRAKLTISPDYAYGATG HPGIIPPHATLVFDVELLKLE**
Frame 2: AVAVGPRRPSRRPPAQRPPPPWECRWKPSPQETGAPSPSAARPAWCTTPGCLKM ERNLIPPGTETSPLSLC
Frame 3: RSLLVHAARRAARPLSVRRRHGSAGGNHLPRRRAHLPQARPDLRGALHRDA

Brain cDNA Library – Round 7 – Plaque G1

TATGTGTGTTCGAGCCGGAC**GAATTCA**AGCAGACAGGACCAGAAGAACTAAAGAAACTTTCAGAAGTGGGCAG
AAGAGAGCCAGAAGAACTGGGAAAAACACAATTTGGAGAAATAGGAGAAACGAAAAAAACAGGAAATGAGATG
GGAAAGGAATATGAATGAAGCCATCGGTAGAGATGAGGATCAGGAAGCTGGTGTTCAGAGGGATCATGGGATT
TTATTAAACTGGATTTTCAAGCGATTTGTCTGTTATAGGAAAAATGAGGGTTTTACTTCTGCTGCTTTCCATC
ACTATTTTGCCATTAAATAGGTGTCTTTCACTCTTGCAAACCCTGAGTCTGTCACTTTGCCTCTTCACCCCTG
CCCATTCTTGGAAGAGCATCGTGGAAGAAATGAGAAAGGTGCCAAGAAGAAAGGGTTTCAGGAGGGTGAAGAT
ACGGGCGCTACAGGGAACAAGAAGTCACAGGGAGGAGAGGGGAAGGCCTGGGAGGAGAGGTGCTTCCCACACA
CCGGTGACTGACTGAGTCCCTCCAGTGCGCAGGACAGCGGGGAGGCTTTATGGATGGAACAGTGGTGGGGCGG
CCCCAGCTGTCACAGGTAGGAGAATGCCGCAAGCCAGAACCAAGCGAATGCTGGGAGTGAGGCCAAAGATAGG
GGAGTGGCCTGGGAGTCGGGAGGGCAGACAGACTCCGCCTCTGACACCTCTGGGTGGCATAATCAGGTAAATC
AGGCTCTCCAAGCTGGCTTCTCCTCTGAACAATGGGATACCTGCCTCAGGGGTTGCTGCGAGGACTGAGTGCT
TAGCACAGCACTTGGACAGGAAACTCAAAAATCCAACGGAGCTCTCAGTGGTTTCACAGAGGAGGTGGCAGCT
GAGGATCTGGCCTCAGAGGATGTTAGGATTTCACTCCGTTACTCATCTGTCCCTTTGGGCCATTGCCATCCTG
GTCTGTCTCCACAGCTCCTAGAAGGAAAGGTCGGGGCTGAGACTGACTTTCGGCTGGCACAGGTCTGGACGAA
TGAGGCCAGTGAGGCATTTCAGATGCAGAGTTTACCGGGAATG (up primer)

LOCUS NM_031421 2298 bp mRNA linear PRI 25-MAR-2012
DEFINITION **Homo sapiens tetratricopeptide repeat domain 25 (TTC25), mRNA.**
ACCESSION NM_031421 VERSION NM_031421.2 GI:134152711

CDS 90-2108

```
   1 gctaagaaac ggagcttcca caaaccagat agaggttctc cagcttttct ttgattgtct
  61 ctgctttagc gtctctaaat ccggtcacca tgtcggaccc cgaaggcgag accttgcgaa
 121 gcacctttcc ctcttatatg gccgaaggcg agcggctcta cctgtgcggg gaattttcta
 181 aagccgcgca gagcttcagc aacgctcttt accttcagga tggagacaag aactgcctgg
```

```
  //
1861 caagatcagg agaaacaggc aggaagctac tagaagctgg cagaagagag tcaagagaaa
1921 tttataggag gccttcggga gaattagagc aaagactctc aggagaattc agcagacagg
1981 aaccagaaga actaaagaaa ctttcagaag tgggcagaag agagccagaa gaactgggaa
2041 aaacacaatt tggagaaata ggagaaacga aaaaaacagg aaatgagatg gaaaaggaat
2101 atgaatgaag ccatcggtag agatgaggat caggaagctg gtgttcagag ggatcatggg
2161 attttattaa actggatttt caagcgattt gtctgttata ggaaaaatga gggttttact
2221 tctgctgctt tccatcacta ttttgccatt aaataggtgt ctttcactct tgcaaaaaaa
2281 aaaaaaaaaa aaaaaaaa
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SADRNQKN**
Frame 2: QQTGTRRTKETFRSGQKRARRTGKNTIWRNRRNEKNRK
Frame 3: SRQEPEELKKLSEVGRREPEELGKTQFGEIGETKKTGNEMGKEYE

8.3.3.2 Manzamine

8.3.3.2.1 Human colon

Colon cDNA Library – Round 15 – Plaque C4

GTTCTGTCGTCCGTCTGGAC**GAATTCA**GCGTCTCAGCTAATCCAAAGGGAGTTTCAGACCAGTGCAATCAGCA
GAGACATTGATACTGCTGCCAAATTTATTGGTGCAGGTGCTGCAACAGTAGGAGTGGCTGGTTCTGGTGCTGG
TATTGGAACAGTCTTTGGCAGCCTTATCATTGGTTATGCCAGAAACCCTTCGCTGAAGCAGCAGCTGTTCTCA
TATGCTATCCTGGGATTTGCCTTGTCTGAAGCTATGGGTCTCTTTTGTTTGATGGTTGCTTTCTTGATTTTGT
TTGCCATGTAACAAATTACTGCTTGACATGTTGGCATTCATATTAATTACGGATGTAATTCTGTGTATCTTAC
TGTGACTCCGAAAACTGGTATTGGTGTCATGGGAATGTACGTTATTTCCAAAGTCATTTCATTAAAGATGAAA
ACTTTAAAAAAAAAAAAAAAAAAAAAAAAAGTTGGGGGCCCCCCTCAAGAAACTATTTAACCCCTGGGGGCCTC
TAAACGGGTTTGGGGGGGGGTAAA (up primer)

LOCUS NM_001689 2651 bp mRNA linear PRI 25-MAR-2012
DEFINITION **Homo sapiens ATP synthase, H+ transporting, mitochondrial Fo complex, subunit C3 (subunit 9) (ATP5G3), nuclear gene encoding mitochondrial protein, transcript variant 2, mRNA.**
ACCESSION NM_001689 VERSION NM_001689.4 GI:149999603

CDS 181-609
```
   1 cttcatccgg gtgctgcggc gcgaataaga gccggaccgc gcttgcgcat tgagtcccac
  61 tccttcgacc tctgccgcag cccgtgccgc cgccgcctcc tgggaagaga ggaagcggga
 121 gaggagccca cgtcgcctgt cacccaatat ctccagccgc gcagtcccga agagtgtaag
 181 atgttcgcct gcgccaagct cgcctgcacc ccctctctga tccgagctgg atccagagtt
 241 gcatacagac caatttctgc atcagtgtta tctcgaccag aggctagtag gactggagag
 301 ggctctacgg tatttaatgg ggcccagaat ggtgtgtctc agctaatcca aagggagttt
 361 cagaccagtg caatcagcag agacattgat actgctgcca aatttattgg tgcaggtgct
 421 gcaacagtag gagtggctgg ttctggtgct ggtattggaa cagtctttgg cagccttatc
 481 attggttatg ccagaaaccc ttcgctgaag cagcagctgt tctcatatgc tatcctggga
 541 tttgccttgt ctgaagctat gggtctcttt tgtttgatgg ttgctttctt gattttgttt
 601 gccatgtaac aaattactgc ttgacatgtt ggcattcata ttaattacgg atgtaattct
 661 gtgtatctta ctgtgactcc gaaaactgta gtattggtgt catgggaatg tacgttattt
 721 ccaaagtcat ttcattaaag atgaaaactt taatttcttc tgtgatttgt acttacacta
 781 agtttagatt atcacaaaga agaacgtgca ttcaggcaga tgctgtccca ttcagaggaa
 841 gctacagcag ttgctccact gatgaaaaat attccaatgt aatttttatg ggaattcttt
  //
2521 ctgaaactac tatttaatac aaagtacata gtataaaatt ttaaaagtga taggaaaatg
2581 cttagattaa tattcaccta ttccaagtat attttaaata aaaacagtct tggtaaagaa
2641 aaaaaaaaaa a //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SASQLIQREFQTSAISRDIDTAAKFIGAGAATVGVAGSGAGIGTVFGSLIIGYARN PSLKQQLFSYAILGFALSEAMGLFCLMVAFLILFAM**
Frame 2: QRLS
Frame 3: SVSANPKGVSDQCNQQRH

Colon cDNA Library – Round 15 – Plaque D4

CCCGGGAAAATTTTCCCGGGGACCA**GAATTCA**CAGCGTGAACTCTGACAACAGTGGCTGGACTGAAATCCGCC
GGGAAGCCTGGGTCTCCTCTAGCTTATTTGGTGTCTCCAGAGCTGTCCAGGAATTTGGTCTTGCCCGATTCAA
AAGCAACGTGACCAAGACTATGAAGGGTTTTGAATATATCTTGGCTAAGCTGCAAGGCGAGGCCCCTTCCAAA
ACACTTGTTGAGACAGCCAAGGAAGCCAAGGAGAAGGCAAAGGAGACGGCACTGGCAGCTACAGAGAAGGCCA
AGGACCTCGCCAGCAAGGCGGCCACCAAGAAGCAGCAGCAGCAGCAACAGTTTGTGTAGCCAGTCTACCACCA
CCACAGCACCCCAGACAGCTAGGCTTAGCCTCTCTGCCCTCCCTTCATTGTACTTTATCATTAAAAATCAAGC

TTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTA (up primer)

LOCUS NM_013237 1287 bp mRNA linear PRI 25-MAR-2012
DEFINITION **Homo sapiens PRELI domain containing 1 (PRELID1), nuclear gene encoding mitochondrial protein, mRNA.**
ACCESSION NM_013237 VERSION NM_013237.2 GI:31543450

CDS 153-812

```
   1 gcggcagctg cttgggcgcg gtgcggtggt gactgagcta cgagcctggc ggcgggtgtg
  61 cgccgagccc cggcccggcc cggccctcgc gtgcctccca ggctccgcac ccctgatgct
 121 gcgcgggtgc tgagcccgct tcggccggga cgatggtgaa gtatttcctg ggccagagcg
 181 tgctccggag ttcctgggac caagtgttcg ccgccttctg gcagcggtac ccgaatccct
 241 atagcaaaca tgtcttgacg gaagacatag tacaccggga ggtgacccct gaccagaaac
 301 tgctgtcccg gcgactcctg accaagacca acaggatgcc acgctgggcc gagcgactat
 361 ttcctgccaa tgttgctcac tcggtgtacg tcctggagga ctctattgtg gacccacaga
 421 atcagaccat gactaccttc acctggaaca tcaaccacgc ccggctgatg gtggtggagg
 481 aacgatgtgt ttactgtgtg aactctgaca acagtggctg gactgaaatc cgccgggaag
 541 cctgggtctc ctctagctta tttggtgtct ccagagctgt ccaggaattt ggtcttgccc
 601 ggttcaaaag caacgtgacc aagactatga agggttttga atatatcttg gctaagctgc
 661 aaggcgaggc cccttccaaa acacttgttg agacagccaa ggaagccaag gagaaggcaa
 721 aggagacggc actggcagct acagagaagg ccaaggacct cgccagcaag gcggccacca
 781 agaagcagca gcagcagcaa cagtttgtgt agccagtcta ccaccaccac agcaccccag
 841 acagctaggc ttagcccctc tgccctccct tcattgtact ttatcattaa aaatcaactt
 901 ccagccctgt ctgctgtcta cgtggtgggt tgtggggatg cagtttggca tctgcagtac
 961 accaagcaca tgattcatgt ctgagccagg tctgcttatt ctcccattgg gcagctgagg
1021 accgaggcac agaggtgcgg tgacttgccc ggggctccag gtagcctgca ggttaactgg
1081 cggtaagtgc tagactgtaa gcccgacaag ggcagggctt ttggttttgt tctctgatgt
1141 gtctcagtat ctagcacata atagacactc aataaatact tgttgaattc aaaaaaaaaa
1201 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
1261 aaaaaaaaaa aaaaaaaaaa aaaaaa
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SSVNSDNSGWTEIRREAWVSSSLFGVSRAVQEFGLARFKSNVTKTMKGFEYILAKLQ**
GEAPSKTLVETAKEAKEKAKETALAATEKAKDLASKAATKKQQQQQQFV
Frame 2: QA
Frame 3: KREL

Colon cDNA Library – Round 15 – Plaque F4

GTCGGAAGCTACTGCGTAC**GAATTCA**GCACAGGAAGTACATCGTGTCCAGGCGGATCTTCTTTCTGGCACTGG
CCTCTCGGTACCGCCGCTCACCATCGCCGGGCTGGGGCCGAATCCAGGCGGATCTTCTTTCTGGCACTGGCCT
CTCGGTACCGCCGCTCACCATCGCCGGGCTGGGGCCGAATCCAGGGCAGCAACTGGGGCTGGTCATCTGCGTC
CAGGACCACAAAGGTCATAAAGGCACTGTTGATGTGGCGCCGGTGGGTCTCAGCCTCCTGGCGATAGGCCTCC
ACGCACACGCCCACCTCCATGCTGAGGGGAGGGAGTGTGGCTTCCAACCTGGACAGACCTTCCCAAGGGCCAT
GCCTCCTCCCCAACACACACACTCACACACACACATGCACACATGCGCCCACACACATGCATGCACACACACA
CACGTATGCACACACATACACACACATGTATGCACACACATACACACACATGTATGCACACACACACTCACAT
ACACTCCTGCAGTTCTTCCAGCAACGTGGATCAGAGTCCCAGTGACAGGACACCTGGGACCGCTCCCTCCACC
ATTGTTTACAGAGCACCTCTTGTGTGTGTCCGGCCCTGTGAGGATGCGAGAGTGAATAAACCTGTCCAGTCTA
TGCCTCATGGGCTAACAGTCAGGGGAGGAGGCAACCAACAAGAAAGCAAAGGTATAAATATGCCATCAAAAAC
TGCAAAGGGATGTAAGGACATAAACCAGGAGGTGGAGGTGGACGGCCTAATTAGAAAATGTTCATTATGTGCA
TATTTACAAGTGCTGTGGCAGCCTTGCTTCAAACAGCTCAGTTGCTGCTATTCATGTATCTTACTGTCTTTAT
TTCTTTATTCCTTTATTTATTTTTTGAGACAGTCTCATTCTGTCGCCCAGGCTGGAGTACAGTGGTACAATCT
TGGCTCACTGCAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAACGGCTCTTGTTGT
GGAGGGGTATAAA (up primer)

LOCUS NM_147161 3181 bp mRNA linear PRI 29-APR-2012
DEFINITION **Homo sapiens acyl-CoA thioesterase 11 (ACOT11), transcript variant 2, mRNA.**
ACCESSION NM_147161 VERSION NM_147161.3 GI:311078763

Gene in backwards

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: SSTGSTSCPGGSSFWHWPLGTAAHHRRAGAESRRIFFLALASRYRRSPSPGWGRIQGS
NWGWSSASRTTKVIKALLMWRRWVSASWR
Frame 2: QAQEVHRVQADLLSGTGLSVPPLTIAGLGPNPGGSSFWHWPLGTAAHHRRAGAESRAA
TGAGHLRPGPQRS
Frame 3: KHRKYIVSRRIFFLALASRYRRSPSPGWGRIQADLLSGTGLSVPPLTIAGLGPNPGQQ
LGLVICVQDHKGHKGTVDVAPVGLSLLAIGLHAHAHLHAEGRECGFQPGQTFPRAMPP
PQHTHSHTHMHTCAHTHACTHTHVCTHIHTHVCTHIHTHVCTHTLTYTPAVLPATWIR
VPVTGHLGPLPPPLFTEHLLCVSGPVRMRE

BLASTp does not reveal any clear target for frame 1 to frame 3, i.e. matches are less than 50% Max Ident, query overage <44%

Colon cDNA Library – Round 15 – Plaque H4

TTACGGAATCCGGGTAC**GAATTCA**GCGACTGTGGGTACAAGGCAGACAGGGAACACGGTGACCCCTGCACCCA
CCCCAACCCCAGGAAGGGCAGGGGAGCAGCGAGGGCCCAGGAAGACCGGGGCCGGGGAAGGACAGCAACAGGG
TCTGAGTCCCTGGGCAGGGGGGCTCAGGGGCTTGTAAACATTGACCACTCTCAGAAGGGGATGGGGGTGGGAG
AGGGCTCCCCCTATACTTCCTCACAGGGTTAAGGCCTCTCCAGGCTCATGGGCCCCCTAAGTCCAAAGTCTCT
TTAGCTGGGGAGAAGAGACAGGAAGAGTGTCCCCCTTCCCCTGCATTGGCTGCGGCAGGGGTGGGGGGGTTAC
ATTCAGTCACAACAGGAGTCTCCCCCACACCTGGCAAGAGGGAAGGGGAGCTGGGGGCCAGCAGCAGCAAACA
CTCCCCCCGGCCCGTCACCCCTTATTGCTTTAGAGAGAATATTGTCTTTTCAGGGACGCTAACACTGTGGGAA
CCAGGGAGAGGCAGGTGGGGGCCCCCCACCCCGAGGCCAGCTTGATCCCCCAACACTGACCTAGTTGCCACTT
TGGTGCAAAAAAAAAAAAAAAAAAAAATCCAACAACAAAAACCCAAACCCCTTCTTTCTGGGACTAGATGGGG
AAAAAAACGGGCCCTCCCCCATGTCCCCGGGGCAGGAGGGAAAAATACTTTCTTGGAGGGAGGCAGTGGCAAA
CCCTTGCCCTAATACTGCACTCTGAGCAATGAGGTCAAGGGGAGGAGTTGGATGAAAAACCCCAGACGCCATG
AAAGCTGCCATGGCCCTGGAAAAAAAGCTGACCAGGGCCCTTTTTGGATCTTCATGCCCTGGGTTCTGCCCGC
ACGGGACCCCCATCTCTTGTCACTTTCTTGGGAGACTCACTTCCTAGATGAGGAAGTGAAGCTTTGCGGCCGC
ACTTCGAGTAAACTAGTTTACCCCCTTGGGGGCCCTCTAAACAGT (up primer)

LOCUS NM_022574 6329 bp mRNA linear PRI 25-MAR-2012
DEFINITION **Homo sapiens GRB10 interacting GYF protein 1 (GIGYF1), mRNA.**
ACCESSION NM_022574 VERSION NM_022574.4 GI:197245455

Gene in backwards

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: SATVGTRQTGNTVTPAPTPTPGRAGEQRGPRKTGAGEGQQQGLSPWAGGLRGL-TLT
TLRRGWGWERAPPILPHRVKASPGSWAP
Frame 2: QRLWVQGRQGTR
Frame 3: SDCGYKADREHGDPCTHPNPRKGRGAARAQEDRGRGRTATGSESLGRGAQGLVNIDH
SQKGMGVGEGSPYTSSQG

BLASTp does not reveal any clear match for proteins in frame 1 and frame 3

8.3.3.2.2 Human Alzheimer's brain

Alzheimer's Brain cDNA Library – Round 15 – Plaque A1

TCTTTATTCGAGCCGGTAC**GAATTCA**GCGGCGGAGGTGCAGGTCCTGGTGCTTGATGGTCGAGGCCATCTCCT
GGGCCGCCTGGCGGCCATCGTGGCTAAACAGGTACTGCTGGGCCGGAAGGTGGTGGTCGTACGCTGTGAAGGC
ATCAACATTTCTGGCAATTTCTACAGAAACAAGTTGAAGTACCTGGCTTTCCTCCGCAAGCGGATGAACACCA
ACCCTTCCCGAGGCCCCTACCACTTCCGGGCCCCCAGCCGCATCTTCTGGCGGACCGTGCGAGGTATGCTGCC
CCACAAAACCAAGCGAGGCCAGGCCGCTCTGGACCGTCTCAAGGTGTTTGACGGCATCCCACCGCCCTACGAC
AAGAAAAAGCGGATGGTGGTTCCTGCTGCCCTCAAGGTCGTGCGTCTGAAGCCTACAAGAAAGTTTGCCTATC
TGGGGCGCCTGGCTCACGAGGTTGGCTGGAAGTACCAGGCAGTGACAGCCACCCTGGAGGAGAAGAGGAAAGA
GAAAGCCAAGATCCACTACCGGAAGAAGAAACAGCTCATGAGGCTACGGAAACAGGCCGAGAAGAACGTGGAG
AAGAAAATTGACAAATACACAGAGGTCCTCAAGACCCACGGACTCCTGGTCTGAGCCCAATAAAGACTGTTAA
TTCCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGTTTGGGGCCCCCCCCCAAAAAATTATTAACC
CCTTGGGGCCCCTAAAGGGTTTTGGGGGGGGGGTTAA (up primer)

LOCUS NM_012423 1142 bp mRNA linear PRI 21-APR-2012
DEFINITION **Homo sapiens ribosomal protein L13a (RPL13A), mRNA.**
ACCESSION NM_012423 VERSION NM_012423.2 GI:14591905

CDS 23-634
1 cttttccaag cggctgccga ag**atggcgga ggtgcaggtc ctggtgcttg atggtcgagg**
61 **ccatctcctg ggccgcctgg cggccatcgt ggctaaacag gtactgctgg gccggaaggt**
121 **ggtggtcgta cgctgtgaag gcatcaacat ttctggcaat ttctacagaa acaagttgaa**
181 **gtacctggct ttcctccgca agcggatgaa caccaaccct tcccgaggcc cctaccactt**
241 **ccgggccccc agccgcatct tctggcggac cgtgcgaggt atgctgcccc acaaaaccaa**
301 **gcgaggccag gccgctctgg accgtctcaa ggtgtttgac ggcatcccac cgccctacga**
361 **caagaaaaag cggatggtgg ttcctgctgc cctcaaggtc gtgcgtctga agcctacaag**
421 **aaagtttgcc tatctggggc gcctggctca cgaggttggc tggaagtacc aggcagtgac**
481 **agccaccctg gaggagaaga ggaaagagaa agccaagatc cactaccgga agaagaaaca**
541 **gctcatgagg ctacggaaac aggccgagaa gaacgtggag aagaaaattg acaaatacac**
601 **agaggtcctc aagacccacg gactcctggt ctga**gcccaa taaagactgt taattcctca
661 tgcgttgcct gcccttcctc cattgttgcc ctggaatgta cgggacccag gggcagcagc
721 agtccaggtg ccacaggcag ccctgggaca taggaagctg ggagcaagga aagggtctta
//
1081 ctcctcacct gtattttgta atcagaaata aattgctttt aaagaaaaaa aaaaaaaaaa
1141 aa

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SAAEVQVLVLDGRGHLLGRLAAIVAKQVLLGRKVVVVRCEGINISGNFYRNKLKYL**
AFLRKRMNTNPSRGPYHFRAPSRIFWRTVRGMLPHKTKRGQAALDRLKVFDGIPPP
YDKKKRMVVPAALKVVRLKPTRKFAYLGRLAHEVGWKYQAVTATLEEKRKEKAKIH
YRKKKQLMRLRKQAEKNVEKKIDKYTEVLKTHGLLV
Frame 2: QRRRCRSWCLMVEAISWAAWRPSWLNRYCWAGRWWSYAVKASTFLAISTETS
Frame 3: SGGGAGPGA

Alzheimer's Brain cDNA Library – Round 15 – Plaque B3

CTACATTTAAGATCCGGGAC**GAATTCA**GCCTAGAGTGGGGCAGGCTAAGATGCTGATGGAGCTGTTTCCTGGG
TAGGGTCCATCTCTACCATGTCCCACCTGCCCTTTTTGTCTTGTCCTGCTTGCTGTTAGCCTGTGGGGTGCTA
TAGGAGTTTCCATCCCCCAGGCGCTCGGCATGTTACTTCCGGTGTTCACAACAAAGCTCGAATGCCTCCTCTT
TAATATTGTCTGTGCATCTCTCTCCCAATCCCCACAAGTCACACTTTAGCATCCTTTAATTATCTTGATTGCA
CTTATTGCTTTCAGAAATTGTCTTGTTTATTTATTATTTCTCTCCAAGAAACAATATTCTTGGAGAATTTCTT
GGAGAGAAACAATAAACAGGGTCCTCATCTGCTTCTTCACTAGTGTATTTCCAGTGCTAAGAATAGTGCCTGG
TACACAGTAGGTTGTTTGTAATTATTTGTGGAATGAAGGAAGACTTGCGATATGCCTCT (up primer)

LOCUS NM_002738 8014 bp mRNA linear PRI 23-APR-2012
DEFINITION **Homo sapiens protein kinase C, beta (PRKCB), transcript variant 2, mRNA.**
ACCESSION NM_002738 VERSION NM_002738.6 GI:197100031

Gene in backwards

Alzheimer's Brain cDNA Library – Round 15 – Plaque D2

TTACAACTCGAATCGCGAC**GAATTCA**GCAGAAGAGAGGCAATTATCAGAAAATAAAATTTTGTATCAATTTTG
CCAGAAGTGAGATTTGCTTTCATCCTCTTTACATGATGTAGGTTTATGTTTAAGAAAACATATTTAGGCTAAA
GTTGAGGACAGAAAGAGGGTATGTTTGTATTTAAGTTTTACCTTAAAAAAGAGATAAAGATGGACAGGTGGGG
TGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGTGGGCAGATCACCTGAGGTCGGGAGTTCGAGAC
CAGCCTGACCAACATGGAGAAACCCCATCTCTACTAAAAAAAAAAAAGCTGGCGGCCGCACTCGAGTAACTAG
TTAACCCCTTGGGGCCTCTAAACGGGTCTTTGAGGGGTTAAAAC (up primer)

LOCUS NG_012881 516941 bp DNA linear PRI 01-APR-2012
DEFINITION **Homo sapiens SRY (sex determining region Y)-box 6 (SOX6), RefSeqGene on chromosome 11.**
ACCESSION NG_012881 VERSION NG_012881.1 GI:257796312

matched sequence is outside CDS

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: SAEERQLSENKILYQFCQK
Frame 2: QQKRGNYQKIKFCINFARSEICFHPLYMM
Frame 3: SRREAIIRK

Alzheimer's Brain cDNA Library – Round 15 – Plaque F1

TATGTGTGTTCGAGCCGGAC**GAATTCA**GCAGACAGGACCAGAAGAACTAAAGAAACTTTCAGAAGTGGGCAGA
AGAGAGCCAGAAGAACTGGGAAAAACACAATTTGGAGAAATAGGAGAAACGAAAAAAACAGGAAATGAGATGG
GAAAGGAATATGAATGAAGCCATCGGTAGAGATGAGGATCAGGAAGCTGGTGTTCAGAGGGATCATGGGATTT
TATTAAACTGGATTTTCAAGCGATTTGTCTGTTATAGGAAAAATGAGGGTTTTACTTCTGCTGCTTTCCATCA
CTATTTTGCCATTAAATAGGTGTCTTTCACTCTTGCAAACCCTGAGTCTGTCACTTTGCCTCTTCACCCCTGC
CCATTCTTGGAAGAGCATCGTGGAAGAAATGAGAAAGGTGCCAAGAAGAAAGGGTTTCAGGAGGGTGAAGATA
CGGGCGCTACAGGGAACAAGAAGTCACAGGGAGGAGAGGGGAAGGCCTGGGAGGAGAGGTGCTTCCCACACAC
CGGTGACTGACTGAGTCCCTCCAGTGCGCAGGACAGCGGGGAGGCTTTATGGATGGAACAGTGGTGGGGCGGC
CCCAGCTGTCACAGGTAGGAGAATGCCGCAAGCCAGAACCAAGCGAATGCTGGGAGTGAGGCCAAAGATAGGG
GAGTGGCCTGGGAGTCGGGAGGGCAGACAGACTCCGCCTCTGACACCTCTGGGTGGCATAATCAGGTAAATCA
GGCTCTCCAAGCTGGCTTCTCCTCTGAACAATGGGATACCTGCCTCAGGGGTTGCTGCGAGGACTGAGTGCTT
AGCACAGCACTTGGACAGGAAACTCAAAAATCCAACGGAGCTCTCAGTGGTTTCACAGAGGAGGTGGCAGCTG
AGGATCTGGCCTCAGAGGATGTTAGGATTTCACTCCGTTACTCATCTGTCCCTTTGGGCCATTGCCATCCTGG
TCTGTCTCCACAGCTCCTAGAAGGAAAGGTCGGGGCTGAGACTGACTTTCGGCTGGCACAGGTCTGGACGAAT
GAGGCCAGTGAGGCATTTCAGATGCAGAGTTTACCGGGAATG (up primer)

LOCUS NM_031421 2298 bp mRNA linear PRI 25-MAR-2012
DEFINITION **Homo sapiens tetratricopeptide repeat domain 25 (TTC25), mRNA.**
ACCESSION NM_031421 VERSION NM_031421.2 GI:134152711
CDS 90-2108

```
   1 gctaagaaac ggagcttcca caaaccagat agaggttctc cagcttttct ttgattgtct
  61 ctgctttagc gtctctaaat ccggtcacca tgtcggaccc cgaaggcgag accttgcgaa
 121 gcacctttcc ctcttatatg gccgaaggcg agcggctcta cctgtgcggg gaattttcta
 181 aagccgcgca gagcttcagc aacgctcttt accttcagga tggagacaag aactgcctgg
  //
```

1861 **caagatcagg agaaacaggc aggaagctac tagaagctgg cagaagagag tcaagagaaa**
1921 **tttataggag gccttcggga gaattagagc aaagactctc aggagaattc agcagacagg**
1981 **aaccagaaga actaaagaaa ctttcagaag tgggcagaag agagccagaa gaactgggaa**
2041 **aaacacaatt tggagaaata ggagaaacga aaaaaacagg aaatgagatg gaaaaggaat**
2101 **atgaatga**ag ccatcggtag agatgaggat caggaagctg gtgttcagag ggatcatggg
2161 attttattaa actggatttt caagcgattt gtctgttata ggaaaaatga gggttttact
2221 tctgctgctt tccatcacta ttttgccatt aaataggtgt ctttcactct tgcaaaaaaa
2281 aaaaaaaaaa aaaaaaaa

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: SADRNQKN
Frame 2: QQTGTRRTKETFRSGQKRARRTGKNTIWRNRRNEKNRK
Frame 3: SRQEPEELKKLSEVGRREPEELGKTQFGEIGETKKTGNEMGKEYE

8.3.3.2.3 Human colon tumour

Colon Tumour cDNA Library – Round 9 – Plaque H7

ATTTCGGGGTCGGGGAC**GAATTCA**AGCCTGTCTGTCACTACACTGGGGATCCCAATTCCACATAAGCACTTTT
GGAAGAAAACAGCCAAAGTTGGCCTAAAATTGGCGCTGGAATTTGGGCTGGGAAAAATCTTGTGGTTATTTCC
TTTAAAAAGGAACAAAACTTTAGTATTTAATTAGTTGATTTATTTAATGTAATTTCAAACAATTAAATTATGA
ATAATGCAATGTACAGTAGAATCACGTTTTGATTTTATTAACACTGACCAAGTTTAACTCCATATGAAGTGTA
AGCTTGATATCGTTTATGATGTCTATCAACTGTACCAAAAGTAAAACATTTAAAAACAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAGGTTTGGGGCCCCCCCCAAAAAAATTTTAACCCCCTGGGGGCCCTAAAA
GGGGTTTGGGGGGGTAAAA (up primer)

Matches with 2 transcript variants (1 + 2) of the following gene:

LOCUS NM_025109 3818 bp mRNA linear PRI 15-AUG-2011
DEFINITION **Homo sapiens myosin XIX (MYO19), transcript variant 1, mRNA.**
ACCESSION NM_025109 VERSION NM_025109.5 GI:284005041

1 gcacttccgg gccggccgat tccgcgcgtc tccgcccatc atggcgctcg tgccagtgtc
61 ccgtgggcta cccactcctg tcatgtcctt tgggctggtg caggccagat cttctagggc
//
901 accatacccc agcccatggt cccatgacta ggtggatagt actccttgta cctcctgcaa
961 cccagaaccc tggctgacca ctttgaagga gg**atgctcca gcaggtcaat ggccacaatc**
1021 **cggggtctga tggccaagcc agggagtacc tcagagaaga cctgcaggag ttcctgggtg**
1081 **gggaggtcct gctgtacaaa ctggatgacc tcaccagggt gaatcctgtg acactagaga**
//
3181 **cactgcggta tgctgacatc tgccctgaac cttcacccta cagcattaca ggctttaatc**
3241 **agattctgct ggaaagacac aggctgatcc acgtgacctc ttctgccttc actgggctgg**
3301 **ggtga**tcctt ggtgcctttg tttccacaag gccttttcct gccccctgcc ttgccaaaga
3361 catttaatca gcacacagct gccagactat tcccacagtg ctccaaatgc acatgaacaa
3421 cagtgacggc tccagccttc gacccagagc cccgtgccca gtgcgtcagt gggcctgggg
3481 ttccaggcta catcaagcac tgatggtgtc agggctggta gttaccaaat cagggttaag
3541 aaacatcagg gccacatttc actaccttca cagatcaaac tcagcagcag tcatgactgt
3601 ctgtcactac actggggatc ccaattccac ataagcactt ttggaagaaa acagccaaag
3661 ttggcctaaa attggcgctg gaatttgggc tgggaaaaat cttgtggtta tttcctttaa
3721 aaaggaacaa aactttagta tttaattagt tgatttattt aatgtaattt caaacaatta
3781 aattatgaat aatgcaatgt aaaaaaaaaa aaaaaaaa //

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage
Peptide encoded by Phage DNA (**Bold** = correct reading frame):

Frame 1: SLSVTTLGIPIPHKHFWKKTAKVGLKLALEFGLGKILWLFPLKRNKTLVFN
Frame 2: ACLSLHWGSQFHISTFGRKQPKLA
Frame 3: PVCHYTGDPNST
BLASTp results in a few matches (all <40aa), none of interest

LOCUS NM_001163735 4418 bp mRNA linear PRI 15-AUG-2011
DEFINITION **Homo sapiens myosin XIX (MYO19), transcript variant 2, mRNA.**
ACCESSION NM_001163735 VERSION NM_001163735.1 GI:254939536

```
   1 gcacttccgg gccggccgat tccgcgcgtc tccgcccatc atggcgctcg tgccagtgtc
  //
 901 accatacccc agcccatggt cccatgacta ggtggatagt actccttgta cctcctgcaa
 961 cccagaaccc tggctgacca ctttgaagga gg[atgctcca gcaggtcaat ggccacaatc]
1021 [cggggtctga tggccaagcc agggagtacc tcagagaaga cctgcaggag ttcctgggtg]
  //
3841 [agattctgct ggaaagacac aggctgatcc acgtgacctc ttctgccttc actgggctgg]
3901 [ggtga]tcctt ggtgcctttg tttccacaag gccttttcct gccccctgcc ttgccaaaga
3961 catttaatca gcacacagct gccagactat tcccacagtg ctccaaatgc acatgaacaa
  //
4141 aaacatcagg gccacatttc actaccttca cagatcaaac tcagcagcag tcatgactgt
4201 ctgtcactac actggggatc ccaattccac ataagcactt ttggaagaaa acagccaaag
4261 ttggcctaaa attggcgctg gaatttgggc tgggaaaaat cttgtggtta tttcctttaa
4321 aaaggaacaa aactttagta tttaattagt tgatttattt aatgtaattt caaacaatta
4381 aattatgaat aatgcaatgt aaaaaaaaaa aaaaaaaa //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame):

Frame 1: SLSVTTLGIPIPHKHFWKKTAKVGLKLALEFGLGKILWLFPLKRNKTLVFN
Frame 2: ACLSLHWGSQFHISTFGRKQPKLA
Frame 3: PVCHYTGDPNST

Colon Tumour cDNA Library – Round 9 – Plaque A8

TATCTGGGCCCGCGAC**GAATTCA**CTTCTCTGGTTATGACAATGACCGACCAGGCAATTTGGTGTATCGCTTCT
GTGACGTCAAAGACGAGACCTATGACTTGCTCTACCAGCAATGCGATGCCCAGCCAGGGGCCAGCGGGTCTGG
GGTCTATGTGAGGATGTGGAAGAGACAGCAGCAGAAGTGGGAGCGAAAAATTATTGGCATTTTTTCAGGGCAC
CAGTGGGTGGACATGAATGGTTCCCCACAGGATTTCAACGTGGCTGTCAGAATCACTCCTCTCAAATATGCCC
AGATTTGCTATTGGATTAAAGGAAACTACCTGGATTGTAGGGAGGGGTGACACAGTGTTCCCTCCTGGCAGCA
ATTAAGGGTCTTCATGTTCTTATTTTAGGAGAGGCCAAATTGTTTTTTGTCATTGGCGTGCACACGTGTGTGT
GTGTGTGTGTGTGTGTGTAAGGTGTCTTATAATCTTTTACCTATTTCTTACAATTGCAAGATGACTGGCTTTA
CTATTTGAAAACTGGTTTGTGTATCATATCATATATCATTTAAGCAGTTTGAAGGCATACTTTTGCATAGAAA
TAAAAAAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTCTGGGAGGGGGT
TA (up primer)

LOCUS NM_007173 3806 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens protease, serine, 23 (PRSS23), mRNA.** ACCESSION
NM_007173 VERSION NM_007173.4 GI:122056696

```
   1 gcggcttccc cgaggccgga ggcggggcgg gcgggcctcg ggtggcgcgg ggggcggacc
  61 cgccagctgc ctgcgctgct cgccagcttg ctcgcactcg gctgtgcggc ggggcaggca
 121 tgggagccgc gcgctctctc ccggcgccca cacctgtctg agcggcgcag cgagccgcgg
 //
 841 tccacttcag ccatgcccga gcagatgaaa tttcagtgga tccgggtgaa acgcacccat
 901 gtgcccaagg gttggatcaa gggcaatgcc aatgacatcg gcatggatta tgattatgcc
 961 [ctcctggaac tcaaaaagcc ccacaagaga aaatttatga agattggggt gagccctcct]
1021 [gctaagcagc tgccaggggg cagaattcac ttctctggtt atgacaatga ccgaccaggc]
1081 [aatttggtgt atcgcttctg tgacgtcaaa gacgagacct atgacttgct ctaccagcaa]
1141 [tgcgatgccc agccaggggc cagcgggtct ggggtctatg tgaggatgtg gaagagacag]
```

1201 **cagcagaagt gggagcgaaa aattattggc atttttcag ggcaccagtg ggtggacatg**
1261 **aatggttccc cacaggattt caacgtggct gtcagaatca ctcctctcaa atatgcccag**
1321 **atttgctatt ggattaaagg aaactacctg gattgtaggg aggggtga**ca cagtgttccc
1381 tcctggcagc aattaagggt cttcatgttc ttatttttagg agaggccaaa ttgttttttg
1441 tcattggcgt gcacacgtgt gtgtgtgtgt gtgtgtgtgt gtaaggtgtc ttataatctt
1501 ttacctattt cttacaattg caagatgact ggctttacta tttgaaaact ggtttgtgta
1561 tcatatcata tatcatttaa gcagtttgaa ggcatacttt tgcatagaaa taaaaaaaat
1621 actgatttgg ggcaatgagg aatatttgac aattaagtta atcttcacgt ttttgcaaac
1681 tttgattttt atttcatctg aacttgtttc aaagatttat attaaatatt tggcatacaa
//
3661 cagaggaaga tgcctctcca ttttccctct ctttatcaga ggttcacatg cctgtctgca
3721 cattaaaagc tctgggaaga cctgttgtaa agggacaagt tgaggttgta aaatctgcat
3781 ttaaataaac atctttgatc acaaaa //

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame):

Frame 1: LLWL
Frame 2: **FSGYDNDRPGNLVYRFCDVKDETYDLLYQQCDAQPGASGSGVYVRMWKRQQQK**
WERKIIGIFSGHQWVDMNGSPQDFNVAVRITPLKYAQICYWIKGNYLDCREG
Frame 3: SLVMTMTDQAIWCIASVTSKTRPMTCSTSNAMPSQGPAGLGSM

BLASTp shows protein (frame 2) matches serine protease protein from pos. 279-383. Active sites are reported at pos. 175 & 246. Hence, match is outside active site.

Colon Tumour cDNA Library – Round 9 – Plaque G8

ACTATCGGGGCCGGGAC**GAATTCA**AGCCAGCCTTCTTGGACAAGGAGTTACTACCTTCTACTTTGAAACCAGT
GAAAGGAACTTCAAGGAAGCCCTTCAGAGAGCTTTGAAATGTGTTTCGTACAACCATTTAGCTGGGGAGCCCA
GAGAGGGCTGCTGCCATGCGATCTCCCTGAAATTGTTTTTTTCAGGCCAGGGATGTGAGAATGTTTCCTGAGG
ACCAGATGTGTGACACTCAGGAGAGAGCCGGCATAAAATAGCCCCCAGAGGTAAAGAAAACCTCAGGACTGAG
GCTGGATACTAAGGGTTAGGAAAGAATTACAGACTTTTCTTGAGACAGGGTCTTTGTCATCCAGGCTGGAGTG
CAGTGGCACAATCAAGGTTCACTGCAGCCTTGACCTCCCAGGCTCAAGCAATCCTTCCACCTCAGCTCCCTGA
GTAGCTGGGACTACAGGCACATACCACCATGCCTGGCTAATTTTTTGTATTTTTAATGGAGAAGAGGTTTCAC
CAGGTTGCCCAGGCTGGTCTCAAACTCTTGGGCTCAAGCCAATGGCACCTCTGAGCCTCCCAAAGTGCTGGGA
TGACAAGCATGAACCTCCGAACCTGGCAGAGTATTTTAAATTGAGATGGGTCCATCAGATACAAGGATAACTA
CCCAGCAGGGAACCTCCAAAGAGGTCATACAAATGCCTGAGACAGAAAAAGTCGATTTAATTATCTGCGGGGC
CCAGGGATCATGAAGCCAGCCAGCCAAGGAAGAACTACCCTATCTCCTCACTCCTCACCTTTTGTACCTCCAG
TCTACACTTTAATCCCACTGAGTTAACTTGCAGCTCACAGCTCCAGAAAAGTATAAGGAGAAAGCTGAAATCA
GAGGAATCGTGCCCTCCACTGCCAGCTTGCAGTTTAAAGCACACCCAAGCTATAATTAAGATTACCTTCTGAC
GATTACATAGGACTGCACGTTTTACCTATTGAAATAAAACTG (up primer)

>50 hits for query coverage between approx. 320 – 580 bp

Perfect match from query start to end:

LOCUS NR_037665 3881 bp RNA linear PRI 18-DEC-2010
DEFINITION **Homo sapiens hypothetical LOC100506548, non-coding RNA.**
ACCESSION NR_037665 VERSION NR_037665.1 GI:315221154

Gene in backwards

Colon Tumour cDNA Library – Round 9 – Plaque B9

TTGGGGGGGGCGCGCAC**GAATTCA**AGCATGAAGTGGGAGAATCGGCATGAAGGCGTCTTCAAGTTCCTGCGCT
CCGAGGCTGTGGCCCAACTATGGGGCCAAAAGAAAAAGAACAGCAACATGACCTACGAGAAGCTGAGCCGGGC

CATGAGGTACTACTACAAACGGGAGATCCTGGAACGGGTGGATGGCCGGCGACTCGTCTACAAGTTTGGCAAA
AACTCAAGCGGCTGGAAGGAGGAAGAGGTTCTCCAGAGTCGGAACTGAGGGTTGGAACTATACCCGGGACCAA
ACTCACGGACCACTCGAGGCCTGCAAACCTTCCTGGGAGGACAGGCAGGCCAGATGGCCCCTCCACTGGGGAA
TGCTCCCAGCTGTGCTGTGGAGAGAAGCTGATGTTTTGGTGTATTGTCAGCCATCGTCCTGGGACTCGGAGAC
TATGGCCTCGCCTCCCCACCCTCCTCTTGGAATTACAAGCCCTGGGGTTTGAAGCTGACTTTATAGCTGCAAG
TGTATCTCCTTTTATCTGGTGCCTCCTCAAACCCAGTCTCAGACACTAAATGCAGACAACACCTTCCTCCTGC
AGACACCTGGACTGAGCCAAGGAGGCCTGGGGAGGCCCTAGGGGAGCACCGTGATGGAGAGGACAGAGCAGGG
GCTCCAGCACCTTCTTTCTGGACTGGCGTTCACCTCCCTGCTCAGTGCTTGGGCTCCACGGGCAGGGGTCAGA
GCACTCCCTAATTTATGTGCTATATAAATATGTCAGATGTACATAGAGATCTATTTTTTCTAAAACATTCCCC
TCCCCCACTCCTCTCCCACAGAGTGCTGGACTGTTCCAGGCCCTCCAGTGGGCTGATGCTGGGACCCTTAGGA
TGGGGCTCCCAGCTCCTTTCTCCTGTGATGGAGGCAGAGACCCTCCAATAAAGTGCCTTCTGGGCTTTTTCTA
ACCTTT (up primer)

Matches with 2 transcript variants (1 + 2), which both encode the same protein.

LOCUS NM_004433 3149 bp mRNA linear PRI 04-JUN-2011
DEFINITION **Homo sapiens E74-like factor 3 (ets domain transcription factor, epithelial-specific) (ELF3), transcript variant 1, mRNA.**
ACCESSION NM_004433 VERSION NM_004433.4 GI:167235022

CDS 156-823
1 ctgagctcag ggaggagctc cctccaggct ctatttagag ccgggtaggg gagcgcagcg
61 gccagatacc tcagcgctac ctggcggaac tggatttctc tcccgcctgc cggcctgcct
121 gccacagccg gactccgcca ctccggtagc ctc**atggctg caacctgtga gattagcaac**
181 **atttttagca actacttcag tgcgatgtac agctcggagg actccaccct ggcctctgtt**
//
961 **cccagaggca cccacctgtg ggagttcatc cgggacatcc tcatccaccc ggagctcaac**
1021 **gagggcctca tgaagtggga gaatcggcat gaaggcgtct tcaagttcct gcgctccgag**
1081 **gctgtggccc aactatgggg ccaaaagaaa aagaacagca acatgaccta cgagaagctg**
1141 **agccgggcca tgaggtacta ctacaaacgg gagatcctgg aacgggtgga tggccggcga**
1201 **ctcgtctaca agtttggcaa aaactcaagc ggctggaagg aggaagaggt tctccagagt**
1261 **cggaactga**g ggttggaact atacccggga ccaaactcac ggaccactcg aggcctgcaa
1321 accttcctgg gaggacaggc aggccagatg gcccctccac tggggaatgc tcccagctgt
1381 gctgtggaga gaagctgatg ttttggtgta ttgtcagcca tcgtcctggg actcggagac
//
1741 taatttatgt gctatataaa tatgtcagat gtacatagag atctattttt tctaaaacat
1801 tcccctcccc actcctctcc cacagagtgc tggactgttc caggccctcc agtgggctga
1861 tgctgggacc cttaggatgg ggctcccagc tcctttctcc tgtgaatgga ggcagagacc
1921 tccaataaag tgccttctgg gctttttcta acctttgtct tagctacctg tgtactgaaa
1981 tttgggcctt tggatcgaat atggtcaaga ggttggaggg gaggaaaatg aaggtctacc
//
3061 tgttgtcttg gaaataccaa tcagattgtt ggctgaagtg atgtggataa agaagggatc
3121 ttagaaaaac taaaaaaaaa aaaaaaaaa //

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SMKWENRHEGVFKFLRSEAVAQLWGQKKKNSNMTYEKLSRAMRYYYKREI LERVDGRRLVYKFGKNSSGWKEEEVLQSRN**
Frame 2: A
Frame 3: HEVGESA

Colon Tumour cDNA Library – Round 15 - Plaque B7

CACTTGAAGTCCGAGCAC**GAATTCA**AGCGTGACATCCGAGGTGCCTTTCTCCAAAAGGTATTTGAAATATCTC
ACCAAAAAATATTTGAAGAAGAATAATCTACGTGACTGGTTGCGCGTAGTTGCTAACAGCAAAGAGAGTTACG
AATTACGTTACTTCCAGATTAACCAGGACGAAGAAGAGGAGGAAGACGAGGATTAAATTTCATTTATCTGGAA

AATTTTGTATGAGTTCTTGAATAAAACTTGGGAACCAAAAAAAAAAAAAAAAAAAAAAAAAAGTTGGGGGCCC
CCCCCAAAAAATTTTTAACCCCTTGGGGGCCTTAAAAGGGGTTGGGGGGGTTAAA (up primer)

LOCUS NM_000983 2099 bp mRNA linear PRI 28-APR-2012
DEFINITION **Homo sapiens ribosomal protein L22 (RPL22), mRNA.** ACCESSION
NM_000983 VERSION NM_000983.3 GI:48255919

CDS 47-433
```
   1 gcgtctgcgt agttcgctca cctccctttc taactccgct gccgcatggc tcctgtgaa
  61 aaagcttgtg gtgaaggggg gcaaaaaaaa gaagcaagtt ctgaagttca ctcttgattg
 121 cacccaccct gtagaagatg gaatcatgga tgctgccaat tttgagcagt ttttgcaaga
 181 aaggatcaaa gtgaacggaa aagctgggaa ccttggtgga ggggtggtga ccatcgaaag
 241 gagcaagagc aagatcaccg tgacatccga ggtgcctttc tccaaaaggt atttgaaata
 301 tctcaccaaa aaatatttga agaagaataa tctacgtgac tggttgcgcg tagttgctaa
 361 cagcaaagag agttacgaat tacgttactt ccagattaac caggacgaag aagaggagga
 421 agacgaggat taaatttcat ttatctggaa aattttgtat gagttcttga ataaaacttg
 481 ggaaccaaaa tggtggttta tccttgtatc tctgcagtgt ggattgaaca gaaaattgga
 541 aatcatagtc aaagggcttc ccttggttcg ccactcattt atttgtaact tgacttcttt
 601 ttttttctgc ttaaaaattt caattctcgt ggtaatacca gagtagaagg agagggtgac
  //
1981 ataaaattga aactcactgg ttttctatgt ctttttgaac tcttgtaatc gagttttgat
2041 catattttct attaaagtgg ctaacacctg gctactctta ctgtaaaaaa aaaaaaaaa
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SSVTSEVPFSKRYLKYLTKKYLKKNNLRDWLRVVANSKESYELRYFQINQDEEEEEDED**
Frame 2: QA
Frame 3: KRDIRGAFLQKVFEISHQKIFEEE

Colon Tumour cDNA Library – Round 15 – Plaque C7

TGTTTTGGAGTCCGAGAC**GAATTCA**AGCATCAAAAGCCTTGACAAATTGAAGGAGGTGAAGGAGTTTTTGGGT
GAGAACATATCCAACTTTCTTTCCTTAGCTGGCAATACTTACCAACTCACACGAGGCATTGGGAAGGACATCC
GTGCCCTCAGACGAGCCAGAGCCAATCTTCAGTCAGTACCGCATGCCTCAGCCTCACGCCCCCGGGTCACTGA
GCCAATCTCAGCTGAAAGCGGTGAACAGGTGGAGAGAGTTAAGCTTGCGGCCGCACTCGAGTAACTAGTTAAC
CCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTA (up primer)

LOCUS NM_001136541 2831 bp mRNA linear PRI 23-APR-2012
DEFINITION **Homo sapiens apolipoprotein L, 1 (APOL1), transcript variant 4, mRNA.**
ACCESSION NM_001136541 VERSION NM_001136541.1 GI:211938441

CDS 169-1311
```
   1 gactttcact ttccctttcg aattcctcgg tatatcttgg ggactggagg acctgtctgg
  61 ttattataca gacgcataac tggaggtggg atccacacag ctcagaacag ctggatcttg
 121 ctcagtctct gccaggggaa gattccttgg aggaggccct gcagcgacat ggagggagct
 181 gctttgctga gagtctctgt cctctgcatc tgggtgcaac aaaacgttcc aagtgggaca
 241 gatactggag atcctcaaag taagcccctc ggtgactggg ctgctggcac catggaccca
  //
 781 attaccagca gtaccatgga ctacggaaag aagtggtgga cacaagccca agcccacgac
 841 ctggtcatca aaagccttga caaattgaag gaggtgaggg agtttttggg tgagaacata
 901 tccaactttc tttccttagc tggcaatact taccaactca cacgaggcat tgggaaggac
 961 atccgtgccc tcagacgagc cagagccaat cttcagtcag taccgcatgc ctcagcctca
1021 cgccccgggg tcactgagcc aatctcagct gaaagcggtg aacaggtgga gagggttaat
1081 gaacccagca tcctggaaat gagcagagga gtcaagctca cggatgtggc ccctgtaagc
1141 ttctttcttg tgctggatgt agtctacctc gtgtacgaat caaagcactt acatgagggg
```

1201 **gcaaagtcag agacagctga ggagctgaag aaggtggctc aggagctgga ggagaagcta**
1261 **aacattctca acaataatta taagattctg caggcggacc aagaactgtg a**ccacagggc
1321 agggcagcca ccaggagaga tatgcctggc aggggccagg acaaaatgca aacttttttt
1381 tttttctgag acagagtctt gctctgtcgc caagttggag tgcaatggtg cgatctcagc
  //
2701 aaagcagttt agcattggga ggaagctcag atctctagag ctgtcttgtc gccgcccagg
2761 attgacctgt gtgtaagtcc caataaactc acctactcat caagctggaa aaaaaaaaaa
2821 aaaaaaaaaa a

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SSIKSLDKLKEVKEFLGENISNFLSLAGNTYQLTRGIGKDIRALRRARANLQSVPHA SASRPRVTEPISAESGEQVERVKLAAALE**
Frame 2: QASKALTN
Frame 3: KHQKP

Colon Tumour cDNA Library – Round 15 – Plaque G8

TATCGAGACCCGAGAC**GAATTCA**AGCCGACATTCTTACTGCAGACAAGATATAAATAAGGAACATCTTTGGCA
GAAAGTCTCTAAGCTACATTCAAAGATAACTCTTCTAGAGTTAAAAGAGCAACAAACTCTAGGTAGATTGAAG
TCTTTGGAAGCTCTTATAAGGCAGCTAAAGCAGGAAAACTGGCTATCTGAAGAAAACGTCAAGATTATAGAAA
ACCATTTTACAACATATGAAGTCACTATGATATAGAATAACTAGGTTTTAAAACTATGGCTGTTAAATAAGCT
TGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTTTGGAGGGGTTA (up primer)

LOCUS NM_001130475 3306 bp mRNA linear PRI 29-APR-2012
DEFINITION **Homo sapiens THAP domain containing 5 (THAP5), transcript variant 1, mRNA.**
ACCESSION NM_001130475 VERSION NM_001130475.1 GI:194440667

CDS 155-1342
   1 ggcggaagtg agtcgacaga cgaggcggct ttcccggcag aatgctagcg caggcgcagg
  61 ggctcgagag gcctggacct gtggcgcatc ctcagtgagg agggccgccc tgcatccgtc
 121 gccggccccg gtctccaggg gcctcacccg agtc**atgccc cgctattgcg cagcgatttg**
 181 **ttgtaagaac cgccggggac gaaacaataa agaccggaag ctgagttttt atccatttcc**
 241 **tctacatgac aaagaaagac tggaaaagtg gttaaagaat atgaagcgag attcatgggt**
  //
 961 **atctgttatt gccatttttg tacctgctga aaattctaaa ccctcagtta attcttttat**
1021 **atctgcacaa aaagaaacca cggaaatgga agacacagac attgaagact ccttgtataa**
1081 **ggatgtagac tatgggacag aagttttaca aatcgaacat tcttactgca gacaagatat**
1141 **aaataaggaa catctttggc agaaagtctc taagctacat tcaaagataa ctcttctaga**
1201 **gttaaaagag caacaaactc taggtagatt gaagtctttg gaagctctta taaggcagct**
1261 **aaagcaggaa aactggctat ctgaagaaaa cgtcaagatt atagaaaacc attttacaac**
1321 **atatgaagtc actatgatat ag**aataacta ggttttaaaa ctatggctgt taaataagct
1381 ttttccagcc aaaccaaact acatgtaaag tgaactgttt cctgtataaa gttctcatct
1441 taatgaacct atgagagtgt tgtaaagtta tgaactgtat cctattcttg taatacttat
  //
3181 atggggagat gatagagtta ttacagtgtt attaagtatt cctatgttgc attttgcaaa
3241 tacttaattt aatacagatg aaaataatct tgagtcacaa taaagtaatg atatgttaag
3301 tttcaa

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SAEHSYCRQDINKEHLWQKVSKLHSKITLLELKEQQTLGRLKSLEALIRQLKQENWLS EENVKIIENHFTTYEVTMI**
Frame 2: QPNILTADKI
Frame 3: SRTFLLQTRYK

8.3.3.2.4 Human liver tumour

Liver Tumour cDNA Library – Round 9 – Plaque B10

CGGGAACAACCGGGAC**GAATTCA**AGCGTCCAGGCTCAGGTTGGGGTTGGCTTGGTTTCAATAAGGAACGGGGA
CACTTACAAATTGCTGCTTGTCCAAATCAGGATCCACTGCAAGGAACAACAGGCCTTATTCCACTGCTGGGGA
TTGATGTGTGGGAGCACGCTTACTACCTTCAGTATAAAAATGTCAGGCCTGATTATCTAAAAGCTATTTGGAA
TGTAATCAACTGGGAGAATGTAACTGAAAGATACATGGCTTGCAAAAAGTAAACCACGATCGTTATGCTGAGT
ATGTTAAGCTCTTTATGACTGTTTTTGTAGTGGTATAGAGTACTGCAGAATACAGTAAGCTGCTCTATTGTAG
CATTTCTTGATGTTGCTTAGTCACTTATTTCATAAACAACTTAATGTTCTGAATAATTTCTTACTAAACATTT
TGTTATTGGGCAAGTGATTGAAAATAGTAAATGCTTTGTGTGATTGAATCTGATTGGACATTTTCTTCAGAGA
GCTAAATTACAATTGTCATTTATAAAACCATCAAAAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTT
GGGGCCTCTAAACGGGTCTTGAGGGGTTAA (up primer)

LOCUS NM_000636 1593 bp mRNA linear PRI 18-SEP-2011
DEFINITION **Homo sapiens superoxide dismutase 2, mitochondrial (SOD2), nuclear gene encoding mitochondrial protein, transcript variant 1, mRNA.**
ACCESSION NM_000636 VERSION NM_000636.2 GI:67782304

```
   1 gcggtgccct tgcggcgcag ctggggtcgc ggccctgctc cccgcgcttt cttaaggccc
  61 gcgggcggcg caggagcggc actcgtggct gtggtggctt cggcagcggc ttcagcagat
 121 cggcggcatc agcggtagca ccagcactag cagcatgttg agccgggcag tgtgcggcac
 181 cagcaggcag ctggctccgg ttttggggta tctgggctcc aggcagaagc acagcctccc
 //
 481 tggtggagaa cccaaagggg agttgctgga agccatcaaa cgtgactttg gttcctttga
 541 caagtttaag gagaagctga cggctgcatc tgttggtgtc caaggctcag gttggggttg
 601 gcttggtttc aataaggaac ggggacactt acaaattgct gcttgtccaa atcaggatcc
 661 actgcaagga acaacaggcc ttattccact gctggggatt gatgtgtggg agcacgctta
 721 ctaccttcag tataaaaatg tcaggcctga ttatctaaaa gctatttgga atgtaatcaa
 781 ctgggagaat gtaactgaaa gatacatggc ttgcaaaaag taaaccacga tcgttatgct
 841 gagtatgtta agctctttat gactgttttt gtagtggtat agagtactgc agaatacagt
 901 aagctgctct attgtagcat ttcttgatgt tgcttagtca cttatttcat aaacaactta
 961 atgttctgaa taatttctta ctaaacattt tgttattggg caagtgattg aaaatagtaa
1021 atgctttgtg tgattgaatc tgattggaca ttttcttcag agagctaaat tacaattgtc
1081 atttataaaa ccatcaaaaa tattccatcc atatactttg gggacttgta gggatgcctt
1141 tctagtccta ttctattgca gttatagaaa atctagtctt ttgccccagt tacttaaaaa
1201 taaaatatta acactttccc aagggaaaca ctcggctttc tatagaaaat tgcacttttt
1261 gtcgagtaat cctctgcagt gatacttctg gtagatgtca cccagtggtt tttgttaggt
1321 caaatgttcc tgtatagttt ttgcaaatag agctgtatac tgtttaaatg tagcaggtga
1381 actgaactgg ggtttgctca cctgcacagt aaaggcaaac ttcaacagca aaactgcaaa
1441 aaggtggttt ttgcagtagg agaaaggagg atgtttattt gcagggcgcc aagcaaggag
1501 aattgggcag ctcatgcttg agacccaatc tccatgatga cctacaagct agagtattta
1561 aaggcagtgg taaatttcag gaaagcagaa gtt //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage
Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SVQGSGWGWLGFNKERGHLQIAACPNQDPLQGTTGLIPLLGIDVWEHAYYL QYKNVRPDYLKAIWNVINWENVTERYMACKK**

Frame 2: ASKAQVGVGLVSIRNGDTYKLLLVQIRIHCKEQQALFHCWGLMCGSTLTT
FSIKMSGLII
Frame 3: RPRLRLGLAWFQ

Liver Tumour cDNA Library – Round 9 – Plaque C10

GCCGGGGTCTAGAC**GAATTCA**AGCAATGTGCCCAACCTTCATGTCATGAAGGCCATGCAGTCTCTCAAGTCCC
GAGGCTACGTGAAGGAACAGTTTGCCTGGAGACATTTCTACTGGTACCTTACCAATGAGGGTATCCAGTATCT
CCGTGATTACCTTCATCTGCCCCCGGAGATTGTGCCTGCCACCCTACGCCGTAGCCGTCCAGAGACTGGCAGG
CCTCGGCCTAAAGGTCTGGAGGGTGAGCGACCTGCGAGACTCACAAGAGGGGAAGCTGACAGAGATACCTACA
GACGGAGTGCTGTGCCACCTGGTGCCGACAAGAAAGCCGAGGCTGGGGCTGGGTCAGCAACCGAATTCCAGTT
TAGAGGCGGATTTGGTCGTGGACGTGGTCAGCCACCTCAGTAAAATTGGAGAGGATTCTTTTGCATTGAATAA
AGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTTTGGGAGGGGTTAA (up
primer)

Matches with transcript variant 1, 2, 3 & pseudogene 7

Transcript Variant: This variant (3) differs in the 5' UTR compared
to variant 1. Variants 1, 2 and 3 encode the same protein.

LOCUS NM_001204091 666 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens ribosomal protein S10 (RPS10),transcript variant 3, mRNA.**
ACCESSION NM_001204091 VERSION NM_001204091.1 GI:323276699

1 ggcggggggc gggtccacgc cagcccggaa gagacgcagc accgcgcatg ctccttcctt
61 tccagccccg gtaccggacc ctgcagccgc agaggtga**at gttgatgcct aagaagaacc**
121 **ggattgccat ttatgaactc ctttttaagg agggagtcat ggtggccaag aaggatgtcc**
181 **acatgcctaa gcacccggag ctggcagaca agaatgtgcc caaccttcat gtcatgaagg**
241 **ccatgcagtc tctcaagtcc cgaggctacg tgaaggaaca gtttgcctgg agacatttct**
301 **actggtacct taccaatgag ggtatccagt atctccgtga ttaccttcat ctgccccggg**
361 **agattgtgcc tgccacccta cgccgtagcc gtccagagac tggcaggcct cggcctaaag**
421 **gtctggaggg tgagcgacct gcgagactca caagaggggа agctgacaga gatacctaca**
481 **gacggagtgc tgtgccacct ggtgccgaca agaaagccga ggctggggct gggtcagcaa**
541 **ccgaattcca gtttagaggc ggatttggtc gtggacgtgg tcagccacct cagtaa**aatt
601 ggagaggatt cttttgcatt gaataaactt acagccaaaa aaccttaaaa aaaaaaaaaa
661 aaaaaa //

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SNVPNLHVMKAMQSLKSRGYVKEQFAWRHFYWYLTNEGIQYLRDYLHLPPE**
IVPATLRRSRPETGRPRPKGLEGERPARLTRGEADRDTYRRSAVPPGADKK
AEAGAGSATEFQFRGGFGRGRGQPPQ
Frame 2: AMCPTFMS
Frame 3: QCAQPSCHEGHAVSQVPRLREGTVCLETFLLVPYQ

Liver Tumour cDNA Library – Round 9 – Plaque D10

AGGTACGGGGTCTGGGAC**GAATTCA**GCCGGGGGCTTCTCCATCCAGGTCCCTGGAGTTCCTGGTCCCTGGAGC
TCCGCACTTGGCGGCGCAACCTGCGTGAGGCAGCGCGACTCTGGCGACTGGCCGGCCATGCCTTCCCGGGCTG
AGGACTATGAAGTGTTGTACACCATTGGCACAGGCTCCTACGGCCGCTGCCAGAAGATCCGGAGGAAGAGTGA
TGGCAAGATATTAGTTTGGAAAGAACTTGACTATGGCTCCATGACAGAAGCTGAGAAACAGATGCTTGTTTCT
GAAGTGAATTTGCTTCGTGAACTGAAACATCCAAACATCGTTCGTTACTATGATCGGATTATTGACCGGACCA
ATACAACACTGTACATTGTAATGGAATATTGTGAAGGAGGGGATCTGGCTAGTGTAATTACAAAGGGAACCAA
GGAAAGGCAATACTTAGATGAAGAGTTTGTTCTTCGAGTGATGACTCAGTTGACTCTGGCCCTGAAGGAATGC

CACAGACGAAGTGATGGTGGTCATACCGTATTGCATCGGGATCTGAAGCTTGCGGCCGCACTCGAGTAACTAG
TTAACCCCTTGGGGCCTCTAAACGGGTTTGGAGGGGTTA (up primer)

LOCUS NM_001204183 1938 bp mRNA linear PRI 25-SEP-2011
DEFINITION **Homo sapiens NIMA (never in mitosis gene a)-related kinase 2 (NEK2), transcript variant 2, mRNA.**
ACCESSION NM_001204183 VERSION NM_001204183.1 GI:323510689

```
   1 gcgacggtta aacggggccc aaggcagggg tggcgggtca gtgctgctcg ggggcttctc
  61 catccaggtc cctggagttc ctggtccctg gagctccgca cttggcggcg caacctgcgt
 121 gaggcagcgc gactctggcg actggccggc catgccttcc cgggctgagg actatgaagt
 181 gttgtacacc attggcacag gctcctacgg ccgctgccag aagatccgga ggaagagtga
 241 tggcaagata ttagtttgga aagaacttga ctatggctcc atgacagaag ctgagaaaca
 301 gatgcttgtt tctgaagtga atttgcttcg tgaactgaaa catccaaaca tcgttcgtta
 361 ctatgatcgg attattgacc ggaccaatac aacactgtac attgtaatgg aatattgtga
 421 aggaggggat ctggctagtg taattacaaa gggaaccaag gaaaggcaat acttagatga
 481 agagtttgtt cttcgagtga tgactcagtt gactctggcc ctgaaggaat gccacagacg
 541 aagtgatggt ggtcataccg tattgcatcg ggatctgaaa ccagccaatg ttttcctgga
 601 tggcaagcaa aacgtcaagc ttggagactt tgggctagct agaatattaa accatgacac
 //
1201 tctgttgaag aactacagct tgctaaagga acggaagttc ctgtctctgg caagtaatcc
1261 aggtatgaga atcaacttgg tcaacagaag ctggtgctac aaatgaagaa atgtgcacca
1321 gtgttgctcc caaagtggct taggtagccc ttttcattta caaatctcaa attttaagat
1381 ggatttcatt gaatatatgc atttcaatgg aaacaaaatt ctgttatagc aatgatttat
//
1861 ttaaaaatta gggcatttac aaaactattg tatgtttcag atttcagaga gataaactgc
1921 atttcaagta aaaataaa //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **AGGFSIQVPGVPGPWSSALGGATCVRQRDSGDWPAMPSRAEDYEVLYTIGT GSYGRCQKIRRKSDGKILVWKELDYGSMTEAEKQMLVSEVNLLRELKHPNI VRYYDRIIDRTNTTLYIVMEYCEGGDLASVITKGTKERQYLDEEFVLRVMT QLTLALKECHRRSDGGHTVLHRDLKLAAALE**
Frame 2: PGASPSRSLEFLVPGAPHLAAQPA
Frame 3: RGLLHPGPWSSWSLELRTWRRNLREAARLWRLAGHAFPG

Liver Tumour cDNA Library – Round 9 – Plaque E10

GCTAGCTCGGGCTAGAC**GAATTCA**AGCCTCCTTTTTAaGGAGGGAGTCATGGTGGCCAAGAAGGATGTCCACA
TGCCTAAGCACCCGGAGCTGGCAGACAAGAATGTGCCCAACCTTCATGTCATGAAGGCCATGCAGTCTCTCAA
GTCCCGAGGCTACGTGAAGGAACAGTTTGCCTGGAGACATTTCTACTGGTACCTTACCAATGAGGGTATCCAG
TATCTCCGTGATTACCTTCATCTGCCCCCGGAGATTGTGCCTGCCACCCTACGCCGTAGCCGTCCAGAGACTG
GCAGGCCTCGGCCTAAAGGTCTGGAGGGTGAGCGACCTGCGAGACTCACAAGAGGGGAAGCTGACAGAGATAC
CTACAGACGGAGTGCTGTGCCACCTGGTGCCGACAAGAAAGCCGAGGCTGGGGCTGGGTCAGCAACCGAATTC
CAGTTTAGAGGCGGATTTGGTCGTGGACGTGGTCAGCCACCTCAGTAAAATTGGAGAGGATTCTTTTGCATTG
AATAAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTCTGGAAGGGGTTAA
(up primer)

LOCUS NM_001204091 666 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens ribosomal protein S10 (RPS10), transcript variant 3, mRNA.**
ACCESSION NM_001204091 VERSION NM_001204091.1 GI:323276699

```
  1 ggcgggggcc gggtccacgc cagcccggaa gagacgcagc accgcgcatg ctccttcctt
 61 tccagccccg gtaccggacc ctgcagccgc agaggtgaat gttgatgcct aagaagaacc
```

121 **ggattgccat ttatgaactc ctttttaagg agggagtcat ggtggccaag aaggatgtcc**
181 **acatgcctaa gcacccggag ctggcagaca agaatgtgcc caaccttcat gtcatgaagg**
241 **ccatgcagtc tctcaagtcc cgaggctacg tgaaggaaca gtttgcctgg agacatttct**
301 **actggtacct taccaatgag ggtatccagt atctccgtga ttaccttcat ctgcccccgg**
361 **agattgtgcc tgccacccta cgccgtagcc gtccagagac tggcaggcct cggcctaaag**
421 **gtctggaggg tgagcgacct gcgagactca caagagggga agctgacaga gatacctaca**
481 **gacggagtgc tgtgccacct ggtgccgaca agaaagccga ggctggggct gggtcagcaa**
541 **ccgaattcca gtttagaggc ggatttggtc gtggacgtgg tcagccacct cagtaa**aatt
601 ggagaggatt cttttgcatt gaataaactt acagccaaaa aaccttaaaa aaaaaaaaaa
661 aaaaaa //

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SLLFKEGVMVAKKDVHMPKHPELADKNVPNLHVMKAMQSLKSRGYVKEQFAW**
RHFYWYLTNEGIQYLRDYLHLPPEIVPATLRRSRPETGRPRPKGLEGERPAR
LTRGEADRDTYRRSAVPPGADKKAEAGAGSATEFQFRGGFGRGRGQPPQ
Frame 2: ASFLRRESWWPRRMSTCLSTRSWQTRMCPTFMS
Frame 3: PPF

Liver Tumour cDNA Library – Round 9 – Plaque H10

GGGACTCGAGGAC**GAATTCA**AGCGGAAGGTTATAGACATGAAACAAGTAATAAGAATGTACTGGGGCCGGGCA
TGGTGGCTCATGTCTGTAATCCCAGCACTTTGGGAGGCAGAGGCAGGCAGATCACCTGTGGTCAGGAGTTCGA
GACCAGCCTGGCCAACATGGTGAAACCTTGTCTCTACTAAAAATACAAAAATTAACTGGGTGCAGTGGCACAT
GCCTGTAATCCCTGCTACTCGGGAGGCTGAGGCAGGAGAACCACTTGAACCTGGGAGATGGAGGTTGAAGTGA
GCTGAGATTGTGCCACTGCACTCCAGTCTGGGCAAAAGAGTGAGACTCCATCTCAAAAAAAAAAAAAAAAAAA
AAAGAAAGAAGGAAAAATGACATAAACAAATAAGGAGTAAAATTGACAAATGTCAAATAAACAGTGTTAAAAT
TTTTGTATATGAAAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTCTTGG
AGGGGTTAA (up primer)

>50 matches, all matches align from bp 50-60 onwards, majority is in backwards.

Peptide encoded by Phage DNA (**Bold** = correct reading frame):

Frame 1: SGRL
Frame 2: AEGYRHETSNKNVLGPGMVAHVCNPSTLGGRGRQITCGQEFETSLANMVKPCLY
Frame 3: RKVIDMKQVIRMYWGRAWWLMSVIPALWEAEAGRSPVVRSSRPAWPTW

BLASTp shows no significant similarity for proteins from frame 1-3

Liver Tumour cDNA Library – Round 9 – Plaque A11

GGTTGCTGGCAC**GAATTCA**AGCCCAGGGTTCTCTCTGCATTAGTGGGACTGAGGAGCCAGAGAAGACCCTGAG
AGCTAACCCTGAGTTGTGCGGTTCTCTGCACCTGAACGGGAGTCCAAGTAGCTGCATAGCCAGTAGGCCTTCC
TGGGTGGAAGACATTGGGGATAACCTGTACTATGGACACTACCACGGGTTTGGGGACACTGCTGAAAGCATCC
CAGAACTGAACAGTGTGGTCGAGCATTCCAAGTCCGTGAAGGTGCAGGAGCGGTACGACAGTGCCGTGCTGGG
CACCATGCACCTGCACCACGGCTCCTAGAGACGCTGACCTGGCTCTCGGAAACGCAGGAGTCCTTCCTGGTAG
CCAGCTCAGAATACCCATGTAGCAGCAACTTGAACGAATGTCACAACTTGTACTTTTTTTATATACTTCAACT
TTCTGAAAAAGTAAACTTCGACAAGTTCCCAGCAACTGCTTGTTTGTGCATGAGTAGGGCTTACTAAGTGCAT
AGATGTTTCTACAGTGAGGTGTCCTTTTTATAAGGTGCACTTTTGGAGTTTTTCTGATGCCAATCTCAACATT
GTCTTTTTAATACTGTCACCAGATATTGCCATTTTTCTTTTTGTTAAAAGATTATATGATCAAGATAAATTGG
GGTGGTAAATCAGGTGCCTGGTAATTTATCTCTTTGCACATGGGCATCATTTTAAAAAGCTTGCTTCCACTCT
TTTCTGTAGAATTTGACGGAACACAGCTATTTCCCTATGCAAGGTACAGCCTTACAAAGATTTCTGCAGTGAT
TTGTGTGAAGAAGAGAATGTTTGTCTTTTTCAATGAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTT
GGGGCCTCTAAACGGGCTTTGGAGAGGGGTTAA (up primer)

LOCUS NM_017664 2495 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens ankyrin repeat domain 10 (ANKRD10), mRNA.**
ACCESSION NM_017664 VERSION NM_017664.2 GI:70995240

```
   1 gcgcgctccg ctcggcagcc tgtgggacgc gaccgcggcg ctagtctcgt tcctttgtgc
  61 tgcggcggcg gcttctcgag tcctccccga cgcgtcctct aggccagcga gccccgcgct
 121 ctccggtgac ggacc atgtc ggcggcggga gcgggcgcgg gcgtagaggc gggcttctcc
 181 agcgaggagc tgctctcgct ccgtttcccg ctgcaccgcg cctgccgcga cggggacctg
 //
 961 aatggatgtg tcatcaatgg acatttggac ttcccctcca cgaccccgct cagtgggatg
1021 gaaagcagga atggccagtg cttgacagga actaacggaa ttagcagtgg attagcccca
1081 ggacagccgt ttccgagtag ccagggttct ctctgcatta gtgggactga ggagccagag
1141 aagaccctga gagctaaccc tgagttgtgc ggttctctgc acctgaacgg gagtccaagt
1201 agctgcatag ccagtaggcc ttcctgggtg gaagacattg gggataacct gtactatgga
1261 cactaccacg ggtttgggga cactgctgaa agcatcccag aactgaacag tgtggtcgag
1321 cattccaagt ccgtgaaggt gcaggagcgg tacgacagtg ccgtgctggg caccatgcac
1381 ctgcaccacg gctcctagag acgctgacct ggctctcgga aacgcaggag tccttcctgg
1441 tagccagctc agaataccca tgtagcagca acttgaacga atgtcacaac ttgtacgttt
1501 tttatatact tcaactttct gaaaaagtaa acttcgacaa gttcccagca actgcttgtt
1561 tgtgcatgag tagggcttac taagtgcata gatgtttcta cagtgaggtg tccttttat
1621 aaggtgcact tttggagttt ttctgatgcc aatctcaaca ttgtcttttt aatactgtca
1681 ccagatattg ccatttttct ttttgttaaa agattatatg atcaagataa attggggtgg
1741 taaatcaggt gcctggtaat ttatctcttt gcacatgggc atcattttaa aaagcttgct
1801 tccactcttt tctgtagaat ttgacggaac acagctattt ccctatgcaa ggtacagcct
1861 tacaaagatt tctgcagtga tttgtgtgaa gaagagaacg tttgtctttt tcaatgaagc
1921 tttgcagatc accatgtggt tgaaggtttt agttgtggac acagtggtcc ctccttaatg
1981 atgaagatca ctgccttggg cttcatggaa aacaggccca gcctggggct gcgtttggat
//
2401 aaatgtcaaa tgaacaaacc agtgttctaa gagtgttact aacattttgt tctaaaactg
2461 tccttcacaa attgaataaa aaactctcac actca //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: SPGFSLH
Frame 2: **AQGSLCISGTEEPEKTLRANPELCGSLHLNGSPSSCIASRPSWVEDIGDN**
LYYGHYHGFGDTAESIPELNSVVEHSKSVKVQERYDSAVLGTMHLHHGS
Frame 3: PRVLSALVGLRSQRRP

Liver Tumour cDNA Library – Round 9 – Plaque D11

GGCTGGGGTAC**GAATTCA**AGCGTGCAGAACTGCCCTCACCACCCCTGGCCACCCTGGCCTCTTGGGAGGAACA
GGCAGAGAGGTGGCTTCAGATGGCTCTTGGCTGCCACTCTAGGCCTCGGGGCTTATACAATGAGCAGTGGGCT
CTACCTTCCAATAGGAAGTGCAAACTAATTCGAAGTCACACTTCACCAGGAGGGAGAGATGGTCTTGGCTGAA
GGCACTTTAATCAGGGAGCAAACCCAGTGCCGGAGTTTGTCTTCTCGTTACTCTGATAAATCTAGACACCACT
CTGTTGGAGAACCAGCACTGCCGGCAGCTGGGGTTAGTGCACTGACTGAGCGAGGCATGAGGCCTGCTCACTG
GTAGAACTTCTTGTTAGGGCTGTTAGCATGGTCAGCTAGCAGCTGGCATGGGGTGAACTGTTTTCCATAGGCA
GCTTCATATTTCTTGAGCCGGTCCACTATCTTCTGGGCGCCATACAGATCCACAAAGCGGAAAGGCCCTCCCA
GACAAGGCGGGAAGCCAAGCCCAAAGACGGCTCCGATGTCTCCCTCTGCAGGTGTGGCCAAGATCCCCTCTTG
CAGGCACATGACTGCCTCATTCACAAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTA
AACGGGTTTGGGAGGGGTTA (up primer)

LOCUS NM_000182 3048 bp mRNA linear PRI 13-AUG-2011
DEFINITION **Homo sapiens hydroxyacyl-CoA dehydrogenase/3-ketoacyl-CoA thiolase/enoyl-CoA hydratase (trifunctional protein), alpha subunit (HADHA), nuclear gene encoding mitochondrial protein, mRNA.**
ACCESSION NM_000182 VERSION NM_000182.4 GI:105990523

Gene in backwards

Liver Tumour cDNA Library – Round 15 – Plaque A11

TTTTGGAAATCCTGGGAC**GAATTCA**AGCTGGGCTTCCTCCACACAGGCTATGCTGGTACTGCCCCATCGCCCG
CAGTGGTTCAGCCCAGCCCCAAGGCTCCAGGCTCAGGGGCCCTGGCAGGAAGGAAGGAGATGGGAACTTAGGC
TGCGGAACTATGTCCCGGAGGATGAGGACCTGAAGAAGAGGAGGGTGCCCCAGGCCAAACCGGTTGCAGTGGA
GGAGAAGGTGAAGGAGCAGCTGGAGGCCGCCAAGCCCGAGCCCGTCATCGAGGAGGTGGACCTGGCCAACCTC
GCTCCTCGGAAGCCTGACTGGGACCTCAAGAGAGATGTGGCCAAGAAGCTGGAGAAACTAAAAAAGCGGACTC
AGAGGGCCATTGCCGAGCTGATCCGTGAAAGGCTGAAAGGCCAGGAAGACAGCCTAGCCTCTGCAGTGGATGC
TGCCACCGAACAAAAGACCTGTGACTCCGACTGAGGCATGCCCTGCCCCACCACTCGCCCATCAGGCCTGTCC
TGCAGAGGATGGTCTTGGGCAGGGATGGGGGCTAGGCTTGCCATCACCTCCAGTTTGGCTTCTGAGCAGAGAC
TCCCTGCCCATCAAGTCTGAAACCCCCATGGATGAGGTCAGCTCCTTGTCTGCTGGGTGGCCCCTGCCATTCT
GAATGGAGGCAGAACCAGCAACAACTCTGGGCGTGCCTGTGTCTGCACATGTGGATGTACATATGTCTGTATA
TATGTATATATTTTGAACTTTCTAAAAAAAAATCTGGAAATAGAAACAAGTAAAGCTTGCGGCCGCACTCGAG
TAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTCTTTGAGGGGGTTACAACTAGTTACTCGAGTGCGGCGCA
AGCTTTACTTGTTTTCTATTTCAGATTTTTTTTTAGAAAGTTCAAATATATACATATATACGACATATGTACA
TCCACATGTGCAG (up primer)

LOCUS NM_144716 893 bp mRNA linear PRI 31-MAR-2012
DEFINITION **Homo sapiens coiled-coil domain containing 12 (CCDC12), mRNA.**
ACCESSION NM_144716 VERSION NM_144716.3 GI:238231396

CDS 10-549
1 gcctgcgcg**a tgcaagacgg gagaaaagga ggggcgtacg cgggcaagat ggaggcgact**
61 **acggctggtg tgggccggct agaggaagag gcgttgcggc gaaaggaacg gctgaaggcc**
121 **ctacgggaga aaaccgggcg caaggacaag gaagatgggg agccaaagac caagcatctc**
181 **agagaagagg aggaagaagg cgagaagcac agggaactta ggctgcggaa ctatgtcccg**
241 **gaggatgagg acctgaagaa gaggagggtg ccccaggcca aaccggttgc agtggaggag**
301 **aaggtgaagg agcagctgga ggccgccaag cccgagcccg tcatcgagga ggtggacctg**
361 **gccaacctcg ctcctcggaa gcctgactgg gacctcaaga gagatgtggc caagaagctg**
421 **gagaaactaa aaaagcggac tcagagggcc attgccgagc tgatccgtga aaggctgaaa**
481 **ggccaggaag acagcctagc ctctgcagtg gatgctgcca ccgaacaaaa gacctgtgac**
541 **tccgactga**g gcatgccctg ccccaccact cgcccatcag gcctgtcctg cagaggatgg
601 tcttgggcag ggatgggggc taggcttgcc atcacctcca gtttggcttc tgagcagaga
661 ctccctgccc atcaagtctg aaaccccccat ggatgaggtc agctccttgt ctgctgggtg
721 gcccctgcca ttctgaatgg aggcagaacc agcaacaact ctgggcgtgc ctgtgtctgc
781 acatgtggat gtacatatgt ctgtatatat gtatatattt tgaacttttct aaaaaaaaat
841 ctggaaatag aaacaagtaa acccctgtgt gtggcaaaaa aaaaaaaaaa aaa

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SSWASSTQAMLVLPHRPQWFSPAPRLQAQGPWQEGRRWELRLRNYVPEDEDLKKRRV PQAKPVAVEEKVKEQLEAAKPEPVIEEVDLANLAPRKPDWDLKRDVAKKLEKLKKRT QRAIAELIRERLKGQEDSLASAVDAATEQKTCDSD**
Frame 2: QAGLPPHRLCWYCPIARSGSAQPQGSRLRGPGRKEGDGNLGCGTMSRRMRT
Frame 3: KLGFLHTGYAGTAPSPAVVQPSPKAPGSGALAGRKEMGT

Liver Tumour cDNA Library – Round 15 – Plaque B10

GTTCGGAATCCTACAC**GAATTCA**GCGGAATCCCAAAGCCAAGGTTCTGGCTTCATTGGCAACCCACTAGGAAC
TTGAATGTGAGATTCAGCTCTAGTTTGCACCTTGATTTCCCCAGTGAAAAGGGAGAAGAGCCAAAGTATTGTC
TAGGTGCTGAGTTCCCTTTTCTGAGTGGATTTCAAGCATTTGCCAGGGTATGCTGGAGGAGGGACTCCTAAAG
TAGGTGAGATGTGGGACTAGGTAACACTTTAAATACTTTTTTACTCAAAGATTTGCCAAGAACACAGCTTATT

AGTGAAAGAGACCTGTGGATTGAAATGGCTGCAGATAGATCCCGCAGTCCCTTCTGAAAGGAACTGTGTAGCC
AGCAGGGACCTAAACTTTGGCAAAAAGTGAACACAATTCAAATACGAATGAAGGAAGAGCCCTCAGATGGAAC
GAAAGGCGATGTGCAAGGCCATCCCAGAGAGACGGAAGCAGGTGCGATACTAAGCTTGCGGCCGCACTCGAGT
AACTAGTTAACCCCTTGGGGCCTCTAAACGGG (up primer)

ACGTTTAAGGTTACTGTTCTCGAGTGCGGCCGCAGCTTAGTATCGCACCTGCTTCCGTCTCTCTGGGATGGCC
TTGCACATCGCCTTTCGTTCCATCTGAGGGCTCTTCCTTCATTCGTATTTGAATTGTGTTCACTTTTTGCCAA
AGTTTAGGTCCCTGCTGGCTACACAGTTCCTTTCAGAAGGGACTGCGGGATCTATCTGCAGCCATTTCAATCC
ACAGGTCTCTTTCACTAATAAGCTGTGTTCTTGGCAAATCTTTGAGTAAAAAAGTATTTAAAGTGTTACCTAG
TCCCACATCTCACCTACTTTAGGAGTCCCTCCTCCAGCATACCCTGGCAAATGCTTGAAATCCACTCAGAAAA
GGGAACTCAGCACCTAGACAATACTTTGGCTCTTCTCCCTTTTCACTGGGGAAATCAAGGTGCAAACTAGAGC
TGAATCTCACATTCAAGTTCCTAGTGGGTTGCCAATGAAGCCAGAACCTTGGCTTTGGGATTCCGCTTGAATT
CGGATCCCCGAGCATCACACCTGACTGGAATACGACAGCTCCA (down primer)

LOCUS NM_020781 5508 bp mRNA linear PRI 26-MAR-2012
DEFINITION **Homo sapiens zinc finger protein 398 (ZNF398), transcript variant 2, mRNA.**
ACCESSION NM_020781 VERSION NM_020781.3 GI:194328796

CDS 852-2267
```
   1 tcggtcctca gaggaggggc ccgggcaccc tccgtgcggg ccgcagcacg ccggtctttg
  61 agggaccctt ggactcccgg atctgcatgc gagggatgcg gtcgaagacg agatgaggga
 121 gctgacggga ggggaaacgg aggggcaggt gaacatggag agagcaggaa cagggagaga
  //
 781 tctccatcta cttttccact ccagagtggg aaaaattaga agaatggcaa aaggaacttt
 841 acaagaatat c**atgaagggc aactacgagt ctctcatctc catggattat gctataaatc**
 901 **aacctgatgt cttatctcag attcaaccag aaggggaaca taatacagag gaccaggcag**
 961 **ggccagagga aagtgagatt cccacagacc ccagtgaaga gcctggtatt tcaacatcag**
1021 **atattctgtc ttggattaaa caagaagaag agcctcaggt tggggcccca ccggagtcca**
  //
2101 **caggccacaa tggaggctgt gggggtgata gtgacccatc aggtcagcca cccaacccac**
2161 **caggtcccct cataactggg cttgaaactt ctggcctggg tgtcaacact gaaggtctag**
2221 **agaccaacca gtggtatggg gaagggagtg gaggggggagt tttgtaa**atc caaatctctg
2281 tggcttcatg cttgtatatg ctcacagcag ggcacaaaat ccaagagaag gtctgtgagc
2341 cccatccaac acccacagta attattatct ggcacatcaa tgaatttggg gtcctataca
  //
4201 caataataca gaacaagaca tgatgggcaa taccctagca tattgggatt aagagctttc
4261 tagccagaat ctggacctat accatgggaa aatttagggg tgctgggaat cccaaagcca
4321 aggttctggc ttcattggca acccactagg aacttgaatg tgagattcag ctctagtttg
4381 caccttgatt tccccagtga aaagggagaa gagccaaagt attgtctagg tgctgagttc
4441 ccttttctga gtggatttca agcatttgcc agggtatgct ggaggaggga ctcctaaagt
4501 aggtgagatg tgggactagg taacacttta aatacttttt tactcaaaga tttgccaaga
4561 acacagctta ttagtgaaag agacctgtgg attgaaatgg ctgcagatag atcccgcagt
4621 cccttctgaa aggaactgtg tagccagcag ggacctaaac tttggcaaaa agtgaacaca
4681 attcaaatac gaatgaagga agagccctca gatggaacga aaggcgatgt gcaaggccat
4741 cccagagaga cggaagcagg tgcgatacta agcctcagcc caagaataaa agagagttca
4801 actgttactc ttttctttca tcaatcccct atcagttaga actgcaggcc agttccaaag
4861 agtgtttgct taagattcaa aagtgggtaa gtaaaatgga tcagtgggaa ggataaggtg
  //
5341 tgtgcaagtt tcctattagc ggggaaggtg tggggctctg ctgggcttct cttgagcata
5401 tctgagaagt gatgttcatg tctatctcaa ccaggaccag accctgtgac ttactcggta
5461 ataacagtcc acacaagcct cgttgatatt aaaaaaaaaa aaaaaaaa
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: SADPKAKVLASLATH
Frame 2: QRIPKPRFWLHWQPTRNLNVRFSSSLHLDFPSEKGEEPKYCLGAEFPFLSGFQAFAR

VCWRRDS
Frame 3: SGSQSQGSGFIGNPLGT

BLASTp on frame 2 shows matches of less than 14 aa.

Liver Tumour cDNA Library – Round 15 – Plaque D10

CTTTTTCGGAACCGAGAC**GAATTCA**GCGCCCGAGCAGTTCAGTGATGAAGTGGAACCAGCAACACCTGAAGAA
GGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTC
AAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCC
TGATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAG
TGTTAAAAGAAGGCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCAA
TAAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTTTGGAGGGGTTA (up primer)

LOCUS NM_001472 538 bp mRNA linear PRI 24-MAR-2012
DEFINITION **Homo sapiens G antigen 2C (GAGE2C), mRNA.**
ACCESSION NM_001472 VERSION NM_001472.2 GI:98985825

CDS 84-434
1 acgccaggga gctgtgaggc agtgctgtgt ggttcctgcc gtccggactc tttttcctct
61 actgagattc atctgtgtga aat**atgagtt ggcgaggaag atcgacctat cggcctagac**
121 **caagacgcta cgtagagcct cctgaaatga ttgggcctat gcggcccgag cagttcagtg**
181 **atgaagtgga accagcaaca cctgaagaag gggaaccagc aactcaacgt caggatcctg**
241 **cagctgctca ggagggagag gatgagggag catctgcagg tcaagggccg aagcctgaag**
301 **ctcatagcca ggaacagggt cacccacaga ctgggtgtga gtgtgaagat ggtcctgatg**
361 **ggcaggagat ggacccgcca aatccagagg aggtgaaaac gcctgaagaa ggtgaaaagc**
421 **aatcacagtg ttaa**aagaag gcacgttgaa atgatgcagg ctgctcctat gttggaaatt
481 tgttcattaa aattctccca ataaagcttt acagccttct gcaaagaagt cttgcgca

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SAPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHP QTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC**
Frame 2: QRPSSSVMKWNQQHLKKGNQQLNVRILQLLRRERMREHLQVKGRSLKLIARNRVTH RLGVSVKMVLMGRRWTRQIQRR
Frame 3: SARAVQ

Liver Tumour cDNA Library – Round 15 – Plaque D11

GTTCGGAAATCCGAGAC**GAATTCA**GCGCTGTATGACTCCATGAAGGGCAAGGGGACGCTCTCAGCTCTCGGCG
CACGGCCCAGCTTCCTTCAAAATGTCTACTGTTCACGAAATCCTGTGCAAGCTCAGCTTGGAGGGTGATCACT
CTACACCCCCAAGTGCATATGGGTCTGTCAAAGCCTATACTAACTTTGATGCTGAGCGGGATGCTTTGAACAT
TGAAACAGCCATCAAGACCAAAGGTGTGGATGAGGTCACCATTGTCAACATTTTGACCAACCGCAGCAATGCA
CAGAGACAGGATATTGCCTTCGCCTACCAGAGAAGGACCAAAAAGGAACTTGCATCAGCACTGAAGTAGCCTT
ATCTGGCCACCTGGAGACGGTGATTTTGGGCCTATTGAAGACACCTGCTCAGTATGACGCTTCTGAGCTAAAA
GCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTTTGGGAGGGGTTAAACTAGT
TACTCGAGTGCGGCGCAAGCTTTTAGCTCAGAAGCGTCATACTGAGCAGGTGTCTTCAATAGGCCCAAAATCA
CCGTCTCCAGGTGGCCAGATAGGCTACTTCAGTGCTGATGCAGTTCTTTTTGGTCTTCTCTGGTAGGCGAAGG
CATATCTTGTCTCTGTGCATTGCTGCGGTTGGTCAAATGTGACATGGTGACCTCATCACACCTTTGGTCTTGA
TGGCGGTTTCATGTTCAAAGCATCCGCTTAGCACAAAGTTAGTATAGGCTTGACAGACCATATGCACTTGGGG
GGGTGTAGGGTGATCCCTCCAGCTGAGCTTGCACAGGATTCGTTGAAAGTAGACATTTGATGAGCTGGGCGTG
GCAGAGCTGAGAGCGTCCCTGCCTTCAGGGAGTCATACGCCTGGATTCGGATCCGGAGCACCACCTGACTGGA
ATACGACGCTCTCACA (up primer)

LOCUS NM_004039 1563 bp mRNA linear PRI 06-MAY-2012
DEFINITION **Homo sapiens annexin A2 (ANXA2), transcript variant 3, mRNA.**
ACCESSION NM_004039 VERSION NM_004039.2 GI:50845389 KEYWORDS

CDS 56-1075
1 gctcagcatt tggggacgct ctcagctctc ggcgcacggc ccagcttcct tcaaa**atgtc**
61 **tactgttcac gaaatcctgt gcaagctcag cttggagggt gatcactcta cacccccaag**
121 **tgcatatggg tctgtcaaag cctatactaa ctttgatgct gagcgggatg ctttgaacat**
181 **tgaaacagcc atcaagacca aaggtgtgga tgaggtcacc attgtcaaca ttttgaccaa**
241 **ccgcagcaat gcacagagac aggatattgc cttcgcctac cagagaagga ccaaaaagga**
301 **acttgcatca gcactgaagt cagccttatc tggccacctg gagacggtga ttttgggcct**
361 **attgaagaca cctgctcagt atgacgcttc tgagctaaaa gcttccatga aggggctggg**
421 **aaccgacgag gactctctca ttgagatcat ctgctccaga accaaccagg agctgcagga**
481 **aattaacaga gtctacaagg aaatgtacaa gactgatctg gagaaggaca ttatttcgga**
//
1021 **cactaagggc gactaccaga aagcgctgct gtacctgtgt ggtggagatg actgaagccc**
1081 **gacacggcct gagcg**tccag aaatggtgct caccatgctt ccagctaaca ggtctagaaa
1141 accagcttgc gaataacagt ccccgtggcc atccctgtga gggtgacgtt agcattaccc
1201 ccaacctcat tttagttgcc taagcattgc ctggccttcc tgtctagtct ctcctgtaag
//
1441 tgaggctgtc cctgtaggaa gaaagctctg ggactgagct gtacagtatg gttgccccta
1501 tccaagtgtc gctatttaag ttaaatttaa atgaaataaa ataaaataaa atcaaaaaaa
1561 aaa

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SALYDSMKGKGTLSALGARPSFLQNVYCSRNPVQAQLGG**
Frame 2: QRCMTP
Frame 3: SAV

Liver Tumour cDNA Library – Round 15 – Plaque D12

TTTCCGGAACCTCTAC**GAATTCA**GCAGATAAGGAATACGAGAGAGTACAGGGGCTCAGGTCCAGGTGGCAGGG
GATATGCTACCCAACTCAACTGAGCGGGCCATCACTATTGCTGGCATTCCACAATCCATCATTGAGTGTGTCA
AACAGATCTGCGTGGTCATGTTGGAGACTCTCTCCCAGTCCCCCCCGAAGGGCGTGACCATCCCGTACCGGCC
CAAGCCGTCCAGCTCTCCGGTCATCTTTGCAGGTGGTCAGGACAGGTACAGCACAGGCAGCGACAGTGCGAGC
TTTCCCCACACCACCCCGTCCATGTGCCTCAACCCTGACCTGGAGGGACCACCTCTAGAGGCCTATACCATTC
AAGGACAGTATGCCATTCCACAGCCAGATTTGACCAAGCTGCACCAGTTGGCAATGCAACAGTCTCATTTTCC
CATGACGCATGGCAACACCGGATTCAGTGGCATTGAATCCAGCTCTCCAGAGGTGAAAGGCTATTGGGCAGGT
TTGGATGCATCTGCTCAGACTACTTCTCATGAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGG
CCTCTAAACGGGCTT (up primer)

LOCUS NM_005016 3187 bp mRNA linear PRI 29-APR-2012
DEFINITION **Homo sapiens poly(rC) binding protein 2 (PCBP2), transcript variant 1, mRNA.**
ACCESSION NM_005016 VERSION NM_005016.5 GI:193083105

CDS 351-1451
1 cccagaccag cagaggcagc agccggagca gccgcagcct gcgccctctc ccgcccgccc
61 gccctccgcc cgcccgcccg ccctccgccg ccctccaccc gccccggggt ctctttcccc
121 cttcctcctc ctcctcctcc acccccccct cctcctccgc ccgcccgcgg ggcccccctc
181 gccttcccgc ccgcccctat tgttccgccc ccggcctccc gcccttcccc ttcccgcccg
241 ctccccttt cccctcagtc gcctcgcgcc tgcagttttt ggctttcacc cccaaccagt
301 gaccaaagac ttgaccactc aaagtccagc tccccagaac actgctcgac **atggacaccg**
361 **gtgtgattga aggtggatta aatgtcactc tcaccatccg gctacttatg catggaaagg**
421 **aagttggcag tatcatcgga aagaaaggag aatcagttaa gaagatgcgc gaggagagtg**

gtgcacgtat caacatctca gaagggaatt gtcctgagag aattatcact ttggctggac
ccactaatgc catcttcaaa gcctttgcta tgatcattga caaactggaa gaggacataa
gcagctctat gaccaatagc acagctgcca gtagaccccc ggtcaccctg aggctggtgg
tccctgctag tcagtgtggc tctctcattg gaaaaggtgg atgcaagatc aaggaaatac
gagagagtac aggggctcag gtccaggtgg caggggatat gctacccaac tcaactgagc
gggccatcac tattgctggc attccacaat ccatcattga gtgtgtcaaa cagatctgcg
tggtcatgtt ggagactctc tcccagtccc ccccgaaggg cgtgaccatc ccgtaccggc
ccaagccgtc cagctctccg gtcatctttg caggtggtca ggacaggtac agcacaggca
gcgacagtgc gagctttccc cacaccaccc cgtccatgtg cctcaaccct gacctggagg
gaccacctct agaggcctat accattcaag gacagtatgc cattccacag ccagatttga
ccaagctgca ccagttggca atgcaacagt ctcattttcc catgacgcat ggcaacaccg
gattcagtgg cattgaatcc agctctccag aggtgaaagg ctattgggca ggtttggatg
catctgctca gactacttct catgaactca ccattccaaa cgatttgatt ggctgcataa
tcgggcgtca aggcgccaaa atcaatgaga tccgtcagat gtctggggcg cagatcaaaa
ttgcgaaccc agtggaagga tctactgata ggcaggttac catcactgga tctgctgcca
gcattagcct ggctcaatat ctaatcaatg tcaggctttc ctcggagacg ggtggcatgg
ggagcagcta gaacaatgca gattcatcca taatcccttt ctgctgttca ccaccaccca
tgatccatct gtgtagtttc tgaacagtca gcgattccag gttttaaata gtttgtaaat
tttcagtttc tacacacttt atcatccact cgtgattttt taattaaagc gttttaattc
//
ccctcgagcc tgacttacgg ctgggacagc cccatctttc tgttgattat gtggcgcata
tatatatata tatgtatata tatataattt atataaatat ttctctatgt aaaaaaaaaa
aaaaaaa

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SAEKEIRESTGAQVQVAGDMLPNSTERAITIAGIPQSIIECVKQICVVMLETLSQSPP KGVTIPYRPKPSSSPVIFAGGQDRYSTGSDSASFPHTTPSMCLNPDLEGPPLEAYTIQ GQYAIPQPDLTKLHQLAMQQSHFPMTHGNTGFSGIESSSPEVKGYWAGLDASAQTTSH EACGRTRVTS**
Frame 2: QQKRKYERVQGLRSRWQGICYPTQLSGPSLLLAFHNPSLSVSNRSAWSCWRLSPSPPRRA
Frame 3: SRKGNTREYRGSGPGGRGYATQLN

Liver Tumour cDNA Library – Round 15 – Plaque E10

TTTTCGAACCGGAAC**GAATTCA**GCGGACATCGAGATCGCCACCTACAGGAAGCTGCTGGAGGGCGAGGAGAGC
CGGCTGGAGTCTGGGATGCAGAACATGAGTATTCATACGAAGACCACCAGCGGCTATGCAGGTGGTCTGAGCT
CGGCCTATGGGGGCCTCACAAGCCCCGGCCTCAGCTACAGCCTGGGCTCCAGCTTTGGCTCTGGCGCGGGCTC
CAGCTCCTTCAGCCGCACCAGCTCCTCCAGGGCCGTGGTTGTGAAGAAGATCGAGACACGTGATGGGAAGCTG
GTGTCTGAGTCCTCTGACGTCCTGCCCAAGTGAACAGCTGCGGCAGCCCCTCCCAGCCTACCCCTCCTGCGCT
GCCCCAGAGCCTGGGAAGGAGGCCGCTATGCAGGGTAGCACTGGGAACAGGAGACCCACCTGAGGCTCAGCCC
TAGCCCTCAGCCCACCTGGGGAGTTTACTACCTGGGGACCCCCCTTGCCCATGCCTCCAGCTACAAAACAATT
CAATTGCTTTTTTTTTTTGGTCCAAAATAAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCC
TCTAAACGGGCTTGGGGGGGGGTTA (up primer)

LOCUS NM_001256293 1901 bp mRNA linear PRI 06-MAY-2012
DEFINITION **Homo sapiens keratin 8 (KRT8), transcript variant 3, mRNA.**
ACCESSION NM_001256293 VERSION NM_001256293.1 GI:372466576

CDS 199-1650
acaggccttt ccttacctcc ctccatgctg tccacttcct ctgtaaagct ctcaaccctg
tcccctccc cctctctcct gggaaagagc cctcccatgc ctagctgctg ctcttaggga
ccctgtggct aggtgcgcgg atggaaatcc aggatctccg cctggttcgg cccgcctgcc
tccactcctg cctctacc**at gtccatcagg gtgacccaga agtcctacaa ggtgtccacc**
tctggccccc gggccttcag cagccgctcc tacacgagtg ggcccggttc ccgcatcagc
tcctcgagct tctcccgagt gggcagcagc aactttcgcg gtggcctggg cggcggctat

```
  //
1261 gagctggagg ccgccctgca gcgggccaag caggacatgg cgcggcagct gcgtgagtac
1321 caggagctga tgaacgtcaa gctggccctg gacatcgaga tcgccaccta caggaagctg
1381 ctggagggcg aggagagccg gctggagtct gggatgcaga acatgagtat tcatacgaag
1441 accaccagcg gctatgcagg tggtctgagc tcggcctatg ggggcctcac aagccccggc
1501 ctcagctaca gcctgggctc cagctttggc tctggcgcgg gctccagctc cttcagccgc
1561 accagctcct ccagggccgt ggttgtgaag aagatcgaga cacgtgatgg gaagctggtg
1621 tctgagtcct ctgacgtcct gcccaagtga acagctgcgg cagcccctcc cagcctaccc
1681 ctcctgcgct gccccagagc ctgggaagga ggccgctatg cagggtagca ctgggaacag
1741 gagacccacc tgaggctcag ccctagccct cagcccacct ggggagttta ctacctgggg
1801 acccccctg cccatgcctc cagctacaaa acaattcaat tgctttttt ttttggtcca
1861 aaataaaacc tcagctagct ctgccaatgt caaaaaaaaa a
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SADIEIATYRKLLEGEESRLESGMQNMSIHTKTTSGYAGGLSSAYGGLTSPGLSYSL GSSFGSGAGSSSFSRTSSSRAVVVKKIETRDGKLVSESSDVLPK**
Frame 2: QRTSRSPPTGSCWRARRAGWSLGCRT
Frame 3: SGHRDRHLQEAAGGRGEPAGVWDAEHEYSYEDHQRLCRWSELGLWGPHKPRPQLQPG LQLWLWRGLQLLQPHQLLQGRGCEEDRDT

Liver Tumour cDNA Library – Round 15 – Plaque H11

```
TACGGGGCTCGAGACGAATTCAAGCGGAGCAGACCCGACTGCACATGGCTCTTTTGCTGGAAGAAGAGCATGG
CTGCGCAGAGGACTAAAATTTCATCTGGGAAGGCTTCTTTTGACTGTCAGTAGCAGGATGTCACCAGATGAGG
GTGCTATGGGACCACAGCTGTCTTTGTTCCCATTGCAACTCAACCCTGCGGGAGGCCGCCTGCATCCCTGAGA
GCCTTCTGGAGCCTACAGAGGAGACATTGGCCAGCCAAAAGGAAAGGAGTGGCCAGGGTACGACCTGGAGTAG
GGAAGGGAAAAAGTTCCCGGAAAGAAGAGAATTGGATGAGAGGTCTCGGTGGAAATAAAGGTTTTCTGGCATT
GGTCAAGGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGGTTGGGGCCC
CCCCCCAAAAAAAAAAAAACCCCCTGGGGGCCCTAAAAGGGGTTTGGGGGGGTTT (up primer)
```

```
LOCUS       NM_007210     4520 bp    mRNA    linear   PRI 25-MAR-2012
DEFINITION  Homo sapiens UDP-N-acetyl-alpha-D-galactosamine: poly-peptide
N-acetylgalactosaminyltransferase 6 (GalNAc-T6) (GALNT6), mRNA.
ACCESSION   NM_007210   VERSION     NM_007210.3  GI:115298683

CDS 322-2190
   1 ccaggctggt ccctcctgcc ttcgctctct gctggtcgcg gtggcctgta gctcctccct
  61 cggtaggacc cccctccacc ttggcgctta cgagatgaat gaggagccca gacctctgga
 121 acttggaggg ttgttcagat tcccaggctg acaaaaagga gaaacagctt ttacttgcat
 181 ccagggccaa tcatcgcaga cccagacgtc tgcagagggc tgaggcccca gcttgtggcc
 241 accacaacgt atcaagctat ctccagggtt gggctcagga ctcagagctg acgcagctgg
 301 ggtgcccctt ggttctggag gatgaggctc ctccgcagac gccacatgcc cctgcgcctg
 361 gccatggtgg gctgcgcctt tgtgctcttc ctcttcctcc tgcataggga tgtgagcagc
 421 agagaggagg ccacagagaa gccgtggctg aagtccctgg tgagccggaa ggatcacgtc
  //
2041 caggtcccca aggacgagga atgggaattg gcccaggatc agctcatcag gaactcagga
2101 tctggtacct gcctgacatc ccaggacaaa aagccagcca tggccccctg caatcccagt
2161 gacccccatc agttgtggct ctttgtctag gacccagatc atccccagag agagccccca
2221 caagctcctc aggaaacagg attgctgatg tctgggaacc tgatcaccag cttctctgga
2281 ggccgtaaag atggatttct aaacccactg ggtggcaagg caggaccttc ctaatccttg
  //
4261 gaaagcctga ggcggcagga agagtgtgca gagttgagcg tgtctggaag gctgatccac
4321 tgctgggccc acatcaaagc ccccatgggg agcagacccg actgcacatg gctcttttgc
4381 tggaagaaga gcatggctgc gcagaggact aaaatttcat ctgggaaggc ttcttttgac
```

4441 tgtcagtagc aggatgtcac cagatgaggg tgctatggga ccacagctgt ctttgttccc
4501 attgcaactc aacccctgcgg

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SSGADPTAHGSFAGRRAWLRRGLKFHLGRLLLTVSSRMSPDEGAMGPQLSLFPLQL NPAGGRLHP**
Frame 2: QAEQTRLHMALLLEEEHGCAED
Frame 3: KRSRPDCTWLFCWKKSMAAQRTKISSGKASFDCQ

8.3.3.2.5 Human lung tumour

Lung Tumour cDNA Library – Round 15 – Plaque C3

TCCGTACGGATTCTGGGAC**GAATTCA**GCGTACAGACAGGACAGAATGAAACTCCTGCGGCTCTTTGGCCTGAA
AGTTGGGAATGGTTGGGGGAGAGAAGGGCAGCAGCTTATTGGTGGTCTTTTCACCATTGGCAGAAACAGTGAG
AGCTGTGTGGTGCAGAAATCCAGAAATGAGGTGTAGGGAATTTTGCCTGCCTTCCTGCAGACCTGAGCTGGCT
TTGGAATGAGGTTAAAGTGTCAGGGACGTTGCCTGAGCCCAAATGTGTAGTGTGGTCTGGGCAGGCAGACCTT
TAGGTTTTGCTGCTTAGTCCTGAGGAAGTGGCCACTCTTGTGGCAGGTGTAGTATCTGGGGCGAGTGTTGGGG
GTAAAAGCCCACCCTACAGAAAGTGGAACAGCCCGGAGCCTGATGTGAAAGGACCACGGGTGTTGTAAGCTGG
GACACGGAAGCCAAACTGGAATCAAACGCCGACTGTAAATTGTATCTTATAACTTATTAAATAAAGCTTGCGG
CCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTA (up primer)

LOCUS NM_001164319 9395 bp mRNA linear PRI 23-APR-2012
DEFINITION **Homo sapiens filamin B, beta (FLNB), transcript variant 4, mRNA.**
ACCESSION NM_001164319 VERSION NM_001164319.1 GI:256222414

CDS 166-7902
1 gcggccaggg gcgggcggcc gcagagcagc accggccgtg gctccggtag cagcaagttc
61 gaaccccgct cccgctccgc ttcggttctc gctccttcgg cccttgggcc tccaaacacc
121 agtccccggc agctcgttgc gcattgcgct ctccccgcca ccagg**atgcc ggtaaccgag**
181 **aaggatctag ctgaggacgc gccttggaag aagatccagc agaacacgtt cacacgctgg**
241 **tgcaacgagc acctcaagtg cgtgaacaaa cgcatcggca acctgcagac cgacctgagc**
//
7741 **catgggccca ccacccccctg cgaggaggtc tccatgaagc atgtaggcaa ccagcaatac**
7801 **aacgtcacat acgtcgtcaa ggagaggggc gattatgtgc tggctgtgaa gtgggggggag**
7861 **gaacacatcc ctggcagccc ttttcatgtc acagtgcctt aa**aacagttt tctcaaatcc
7921 tggagagagt tcttgtggtt gcttttgttg cttgtttgta attcatttta tacaaagccc
7981 tccagcctgt ttgtggggct gaaaccccat ccctaaaata ttgctgttgt aaaatgcctt
//
8821 caaggtccgt cccatgacat aacactccac acccgcccca gccaacttca tgggtcactt
8881 tttctggaaa ataatgatct gtacagacag gacagaatga aactcctgcg ggtctttggc
8941 ctgaaagttg ggaatggttg ggggagagaa gggcagcagc ttattggtgg tcttttcacc
9001 attggcagaa acagtgagag ctgtgtggtg cagaaatcca gaaatgaggt gtagggaatt
9061 ttgcctgcct tcctgcagac ctgagctggc tttggaatga ggttaaagtg tcagggacgt
9121 tgcctgagcc caaatgtgta gtgtggtctg ggcaggcaga cctttaggtt ttgctgctta
9181 gtcctgagga agtggccact cttgtggcag gtgtagtatc tggggcgagt gttgggggta
9241 aaagcccacc ctacagaaag tggaacagcc cggagcctga tgtgaaagga ccacgggtgt
9301 tgtaagctgg gacacggaag ccaaactgga atcaaacgcc gactgtaaat tgtatcttat
9361 aacttattaa ataaaacatt tgctccgtaa agttg

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SAYRQDRMKLLRLFGLKVGNGWGREGQQLIGGLFTIGRNSESCVVQKSRNEV**
Frame 2: QRTDRTE
Frame 3: SVQTGQNETPAALWPESWEWLGERRAAAYWWSFHHWQKQ

Lung Tumour cDNA Library – Round 15 – Plaque E1

CCCGGGAAAACACCTTTGGGGATC**GAATTCA**AGCAAAAGAGTTAAACATGAAGTTTCTGGAGAGACAGTGGTA
TTTCAAGGGGGACCTCTTGGCAAAACACCAGATGGTCAGGGCCTTTCTACTTACAAGGAACTTTGTTCTCTGG
CAAGTGATCTTAGCCAGCCAGATCTGGTGTATAAATTTATGAATTTAGCCAACCATCATGCAATGTGGAACTC
TAGGAAGGGTGCTGCTTTTGGTTTTAATGTAATTGCTACCAGAGCTGGAGAGCAGCTGGCTCCTTTTCTGCCT
CAGCTAGTTCCTCGACTTTATCGTTACCAGTTTGATCCCAACCTTGGCATTCGACAGGCCATGACAAGTATTT
GGAATGCGTTGGTCACTGACAAATCCATGGTGGATAAATATTTGAAAGAAATTCTTCAAGATTTGGTTAAGAA
CCTTACAAGCAATATGTGGCGAGTTCGAGAATCCAGCTGTTTAGCTTTGAATGATTTATTGAGAGGAAGACCC
TTAGATGACATCATTGATAAACTTCCAGAAATTTGGGAAACGCTTTTTAGAGTACAAGATGATATCAAGGAAT
CTGTACGAAAAGCGGTAGAACTAGCTCTGAAAACTCTGAGCAAGGTCTGTGTGAAAATGTGTGACCCTGCCAA
AGGAGCAGCTGGCCAGAGAACCATCGCTGCCCTTCTGCCTTGCCTTCTGGACAAAGGAATGATGAGCACCGTG
ACGGAAGTTCGAGCCCTCAGCATTAAGCTTGCCGGCCGCTCTCGAGTAACTAGTTAACCCCTTGGGGCCTCTT
AAATCTCGTCTTGAGGGGGTT (up primer)

LOCUS NM_001080398 7391 bp mRNA linear PRI 22-APR-2012
DEFINITION **Homo sapiens KIAA0368 (KIAA0368), mRNA.**
ACCESSION NM_001080398 XM_001129450 VERSION NM_001080398.1
GI:122937210

CDS 1-6054

```
   1 atgacctcac ggacaaagcg aaggcagaaa gttaaaatca gcatatcgaa aataaatcac
  61 tctagttacc gaaaagaaca taggccgttt tgggagcacc agccacctcc acaccagaac
 121 tgggctgtcg ggggccgctg cggcgcaggg gctcgtccgg gactgagtca gcgccggggc
  //
3421 cttgtcagga agctaagtac ccacaaagaa gtgaagtctc atcttaaaga aattcaaagt
3481 gcatttgttt cagttctatc agaaaatgat gaacttagcc aagatgttgc atcaaagggc
3541 cttgggttgg tttatgaact aggcaatgaa caagatcaac aggaattggt ttctacactt
3601 gtggaaacac ttatgactgg caaaagagtt aaacatgaag tttctggaga gacagtggta
3661 tttcaagggg gagctcttgg caaaacacca gatggtcagg gcctttctac ttacaaggaa
3721 ctttgttctc tggcaagtga tcttagccag ccagatctgg tgtataaatt tatgaattta
3781 gccaaccatc atgcaatgtg gaactctagg aagggtgctg cttttggttt taatgtaatt
3841 gctaccagag ctggagagca gctggctcct tttctgcctc agctagttcc tcgactttat
3901 cgttaccagt ttgatcccaa ccttggcatt cgacaggcca tgacaagtat ttggaatgcg
3961 ttggtcactg acaaatccat ggtggataaa tatttgaaag aaattcttca agatttggtt
4021 aagaacctta caagcaatat gtggcgagtt cgagaatcca gctgtttagc tttgaatgat
4081 ttattgagag gaagaccctt agatgacatc attgataaac ttccagaaat ttgggaaacg
4141 ctttttagag tacaagatga tatcaaggaa tctgtacgaa aagcggcaga actagctctg
4201 aaaactctga gcaaggtctg tgtgaaaatg tgtgaccctg ccaaaggagc agctggccag
4261 agaaccatcg ctgcccttct gccttgcctt ctggacaaag gaatgatgag caccgtgacg
4321 gaagttcgag ccctcagcat taacaccctt gtgaagatca gcaaaagtgc aggagccatg
4381 ttgaaaccgc atgcaccaaa actcattcca gctctgctag agtccttaag tgtattggag
4441 ccccaagttc tcaattattt gagcctccgg gcgacagagc aagaaaaggc tgcgatggat
  //
5941 tctgaatgca gagtgctcct aattgagtct ttagctacta tggagccaga cagcagacct
6001 gaactgcagg agaaagcagc gttactgaag aaaacacttg aaaatctgga ataaattaga
6061 aggggaagaa acaaacaagt gccatgttca ttgggggttg aagtggtggt gttctttgaa
6121 aaaccaagtg ggaaaaagta aagattaatc tgtagcatgc atcattcctt ggctgaaata
  //
```

7261 caccaggaat ctccatgttt attatttttc gtggaaactc catgtttatt atttttggtg
7321 gtctctattg accaatgtat ttcaaacttt tttgactgca agtcataata aaaataactt
7381 tttacatcac g

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SSKRVKHEVSGETVVFQGGALGKTPDGQGLSTYKELCSLASDLSQPDLVYKFMNLAN HHAMWNSRKGAAFGFNVIATRAGEQLAPFLPQLVPRLYRYQFDPNLGIRQAMTSIWN ALVTDKSMVDKYLKEILQDLVKNLTSNMWRVRESSCLALNDLLRGRPLDDIIDKLPE IWETLFRVQDDIKESVRKAVELALKTLSKVCVKMCDPAKGAAGQRTIAALLPCLLDK GMMSTVTEVRALSIKLAGRSRVTS**
Frame 2: QAKELNMKFLERQWYFKGELLAKHQMVRAFLLTRNFVLWQVILASQIWCINL
Frame 3: KQKS

Lung Tumour cDNA Library – Round 15 – Plaque G2

TGTCGCAAGATCCCGAGAC**GAATTCA**GCGTACTGTGTCGCCTCTTCCTTTCCTTCCTTGTCGGAGTACACTCA
TCTTTTTCCTTCTTGTCTTTGGAGCGAGATTCCAATTTTTCTTTATCTTTCTTGTCTCTTTCTCGTTCTCTTT
CTAACTCTTTCTCTTTCTCTCGTTGAAGAAGTTTATCTCTGTAGTGTTCTACTTGCTCCTGAAAACTCTGGCC
TGGTTTTTTAGGTCTTTTCCCAGATTCCAATTCATCCTGAAACTTCATAACTTTGAGCTCAATTTCACGAAGT
TTGGCTCGTTTTTTCCTCACTCATTTCAGAGTACTTAGAGAACTTAGACTCAGTCATTTCTTCTTTGATTGGAT
TAGAGTACAAATGATGTTCTTCAGATTTGGAACTTTGAGTATCTTCTTCATCTTCACTTTCTTCTTCTTGATT
CTGATTTTCTTCTTCTTCTGATTCTTCATGCTGGTCAAATAATTCCCATTTAGAAGTTGTAACAGCCTGTGCT
TCCAATTCAGATTCATCCACAGCTTCCCATTTTGATGGGGCAACTTTAAATATAGGCTCATTTTTTTTTGAGT
CTTCAGTTGCATCCAAAGGCACTCCATCAAGATCATCATCAAGACTTTTTATAGGGACTCCATCAAGATCATC
GATGGGAGTAGCATCAATAGGAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACG
GGTCTTTGAGGGGTTA (up primer)

LOCUS NM_001080415 7464 bp mRNA linear PRI 24-MAR-2012
DEFINITION **Homo sapiens U2 snRNP-associated SURP domain containing (U2SURP), mRNA.**
ACCESSION NM_001080415 XM_031553 VERSION NM_001080415.1 GI:122937226

Gene in backwards

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SAYCVASSFPSLSEYTHLFPSCLWSEIPIFLYLSCLFLVLFLTLSLSLVEEVYLCS VLLAPENSGLVF**
Frame 2: QRTVSPLPFLPCRSTLIFFLLVFGARFQFFFIFLVSFSFSF
Frame 3: SVLCRLFLSFLVGVHSSFSFLSLERDSNFSLSFLSLSRSLSNSFSFSR

BLASTp of frame 1 does not show any significant similarities.

8.3.3.2.6 Human breast tumour

Breast Tumour cDNA Library – Round 9 – Plaque C1

TATCCTAGGTCCGGGAC**GAATTCA**AGCAGCCAGGGCAGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCA
CTCCAGCCTGGGTGACAGAGCAAGACTCCGTTTCAAAAAAAAAAAAAGGGTGGGGGGCGGGGCCTCTGGCTGGA
ATCCCCACCCTTTTGGAAGGCCAAGGGGGGAAAATCCCTGGGGGCCGGGATTTCAAAATAACCCGGGCCAAAT
GGGGAAACCCCCCCCCTTAAAAAAAAAAAAAAATTTACTTGGGCGGGCCGCCCCTTGCCGGTATTCATACTTC
CTGGGAATCCGAAGCCGGGAAATTTGTTTGACCCCAGGGGGCAAAGGTTGCATGGACCCAAAATTGCCCATTT

GCCCTCAACCCTGGCCAACAAAAAGAAAACTCTTCCTCCGAAAAAAACAAAAACCAAAAAAA (up primer)

>50 hits covering only query position 30-120

Breast Tumour cDNA Library – Round 9 – Plaque E1

GAATTCCAAGGCCCAGAAGCAAGGCTGCGACTGTGAGTGTCTGGGCGGCGGGCGCATCTCCCACCAGAGTCAG
GACAAGAAGATTCACGTGTACGGCTATTCCATGGCCTATGGTCCTGCCCAGCACGCCATTTCAACTGAGAAAA
TCAAAGCCAAGTACCCCGACTACGAGGTCACCTGGGCTAACGACGGCTACTGAGCACTCCCAGCCCGGGGCCT
GCTGCCTCCAGCAGCCACTTCAGAGCCCCCGCCTTTGCCTGCACTCCTCTTGCAGGGCTGGCCCTGCCTGCTC
CTGCGGCAGCCTCTGGTGACGTGCTGTCCACCAGGCCTTGGAGACAGGCTAGCCTGGCCACAGAATTAAACGT
GTTGCCACAAAAAAAAAAAAAAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGG
GTTTGGGGGGGGGTTAA (up primer)

LOCUS NM_014172 1218 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens phosphohistidine phosphatase 1 (PHPT1), transcript variant 3, mRNA.**
ACCESSION NM_014172 VERSION NM_014172.4 GI:209529715

CDS 627-1005
```
   1 ggctgccgtt gcctccttgt accggtagcg gggttgggga cggaagcctt cggtcggtgg
  61 agaggagaaa gggagaggcc ttcgggcggt ggacggggaa gagagggagt ccttcgggcg
 //
 541 attggtcgtc gctctcgcag agcccgcctc ccgcagtaca agcggccccc gggtcgggtg
 601 ggaggagggg actccgggag gaggaacatg gcggtggcgg acctcgctct cattcctgat
 661 gtggacatcg actccgacgg cgtcttcaag tatgtgctga tccgagtcca ctcggctccc
 721 cgctccgggg ctccggctgc agagagcaag gagatcgtgc gcggctacaa gtgggctgag
 781 taccatgcgg acatctacga caaagtgtcg ggcgacatgc agaagcaagg ctgcgactgt
 841 gagtgtctgg gcggcgggcg catctcccac cagagtcagg acaagaagat tcacgtgtac
 901 ggctattcca tggcctatgg tcctgcccag cacgccattt caactgagaa aatcaaagcc
 961 aagtaccccg actacgaggt cacctgggct aacgacggct actgagcact cccagccccg
1021 ggcctgctgc ctccagcagc cacttcagag cccccgcctt tgcctgcact cctcttgcag
1081 ggctggccct gcctgctcct gcggcagcct ctggtgacgt gctgtccacc aggccttgga
1141 gacaggctag cctggccaca gaattaaacg tgttgccaca cctgccggct tctgaaaaaa
1201 aaaaaaaaaa aaaaaaaa //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **KAQKQGCDCECLGGGRISHQSQDKKIHVYGYSMAYGPAQHAISTEKIKAKYPD YEVTWANDGY**
Frame 2: RPRSKAATVSVWAAGASPTRVRTRRFTCTAIPWPMVLPSTPFQLRKSKPSTPT TRSPGLTTATEHSQPGACCLQQPLQSPRLCLHSSCRAGPACSCGSLW
Frame 3: GPEARLRL

Breast Tumour cDNA Library – Round 9 – Plaque F1

AAGTTCCATCTTCTGCCTCACTGAGAGGTCTCAGAGTACTGAAAATGAGGACAAATACAATGGCCAGGCAAGA
TGGCTCATGTCTGTAATCCCAGCATTTTGGGAGGCTGTGGCAGGAAGATCACTTGAGCTCAGGAGTTCGAGAC
CAGCCTGGAAACATGGCGAAACCCATCTTTACAAAAATAAAAATAAAAAAATCAGTCAGGCATGGTGGCAAAT
GCCTGTAGTCCCAGCAACTTGGGAGGCTGACGTGGGAGGATCGCTTGAGCCCAGGAGGTCGAGGCTGTCACCA
AGGTGGGTGACAGAGTGAGATCCTGTCTCAAACAAAACAAAACAAACAAAAAACCTTATTAAAAACAAAACAG
GCCAGGCACAGTGGCTCACTTTTATAATCCCAGCACTCTGGGAGGCAGAGGCGGGCAGATTGGTTGAGCTCAG
GAGTTCGAGACCAGCCTGGGCAACATGGCGAAACCCCATCTCTATAAAAATACAAAAATTAGTTGAGTGTGTT
GGCTAGTACCTCTAGTCTCAGCTACTTGGGAGGCTGAGGCAGGAGGATCAATTGAGCCTGTGAGGTGGAGGTT

GCAGTGAGCCGAGAATGTACCAGCCTGGGTGACAGAGGGAGACCGTGTCTCAAAACAAAGCAAAACTAACATC
AACACAGGCTGGTTTGATCCCAGATTAGCAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCC
TCTAAACGGGCTTGG (up primer)

About 50 hits, max identity 82%, too many fragments, no consecutive alignment

Breast Tumour cDNA Library – Round 9 – Plaque H1

AAGTCAAGGCAGCCCTCTTCTCTCCCCGAAGGCCTGCCAGCACCCCTGGAGAAGCGCGTTAAGGAGCTGGCTC
AGCATAGAGGCGCAGACTCGGGAGCTGAGCAGTCAGGTCCGCTCTGGGGTGTATGCCTATCTTGCGTCATTCC
TGCCCTGCAGCAAGGATGCCCTGCTCAAGCGTGCTCGGAAACTTCACCTCTGTGAACAGGGGGGCCGTCTGAA
GGAGCCTCTCCAGAAGCTCAAGGAAGCCATTGGCAGGGCGATGCCAGAGCAGATGGCCAAGTACCAGGACGAA
TGCCAGGCACACACGCAGGCAAAGGTTGCCAAGATGCTGGAAGAGGAGAAAGACAAGGAGCAGAGGGACCGGA
TTTGTTCGGATGAGGAAGAAGATGAAGAAAAAGGGGGCAGGAGGATAATGGGACCTCGGAAGAAGTTCCAATG
GAATGATGAGATCAGGGAGCTACTGTGCCAGGTGGTGAAGATCAAACTGGAGAGCCAGGACCTGGAGAGGAAC
AACAAAGCCCAGGCTTGGGAGGACTGTGTGAAGGGCTTTCTGGATGCGGAAGTCAAGCCCCTCTGGCCCAAAG
GCTGGATGCAGGCCAGAACTCTGTTTAAGGAGAGCAGACGAGGCCATGGGCACCTGACTTCAAGCTTGCGGCC
GCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTCTTTGAGGGGGTTA (up primer)

LOCUS NM_001079514 5982 bp mRNA linear PRI 25-SEP-2011
DEFINITION **Homo sapiens ubinuclein 1 (UBN1), transcript variant 2, mRNA.**
ACCESSION NM_001079514 VERSION NM_001079514.1 GI:118572602

Phage cDNA insert misses a part of the CDS sequence in between 1449-1521, i.e. 71bp. Hence, second part is out of frame.

```
   1 gtcccgcgcc gcaagttctc ctggggcgac cggaggctgc tcccccgacc ccctggtgtc
  61 cccggagtgg ctgcgcggac gtcgagttgg catttcttcg cttcctcctg ggccggcgcg
 121 cggcgcggca caccaggctc ccctgggctc ggggacccgg ccatgggccg aggcgcgggc
 181 cgcccgcccg ctgggagcca cggcttagca gccgaccgct agctgcgccg ccgcccgggg
 241 accggcatga ggaccgccgc ggggggacgt ctgcggcccg cgtcggcgct ggggacaaag
 301 aagccatgca gtgacacccg ctaagacttg ttggtagcca tgtcggagcc ccacagggtc
 361 cagttcacct ctctcccagg ttccctgaat cctgcgtttt tgaagaagtc ccggaaggag
 //
1321 cccttccgag atatggatga tggaagtgat tcccttgggg tgggattgga ccaggaattc
1381 aggcagccct cttctctccc cgaaggcctg ccagcacccc tggagaagcg cgttaaggag
1441 ctggctcagg ctgccagagc tgctgagggg gagagcagac agaagttctt cacccaggat
1501 attaatggaa tcctattaga catagaggcg cagactcggg agctgagcag tcaggtccgc
1561 tctggggtgt atgcctatct tgcgtcattc ctgccctgca gcaaggatgc cctgctcaag
1621 cgtgctcgga aacttcacct ctatgaacag gggggccgtc tgaaggagcc tctccagaag
1681 ctcaaggaag ccattggcag ggcgatgcca gagcagatgg ccaagtacca ggacgaatgc
1741 caggcacaca cgcaggcaaa ggttgccaag atgctggaag aggagaaaga caaggagcag
1801 agggaccgga tttgttcgga tgaggaagaa gatgaagaaa aaggggggcag gaggataatg
1861 ggacctcgga agaagttcca atggaatgat gagatcaggg agctactgtg ccaggtggtg
1921 aagatcaaac tggagagcca ggacctggag aggaacaaca aagcccaggc ttgggaggac
1981 tgtgtgaagg gctttctgga tgcggaagtc aagcccctct ggcccaaagg ctggatgcag
2041 gccagaactc tgtttaagga gagcagacga ggccatgggc acctgacttc aatcctggcc
2101 aagaagaaag ttatggcccc ttctaaaatc aaggtgaagg aatcgtctac gaagcctgat
2161 aaaaaggttt ctgtcccatc aggacagatt ggtggcccca ttgctttgcc ctcagatcac
//
3601 cacagtcttc tggctggctt gcactccagc ccgccccatg cagcgcctct cccacacgct
3661 gcggtgccca cccatatccc gcagagtctg ccaggtgctt ctcagcttca cgggaaaggg
3721 cctgctgtac cacggaaatt gtgaccgctt cagaggcaag gcttgccact tgggtctggg
3781 tggaatcaga acgtgcaggt ctcccaggat gtacactcac tgcgcccttt ctgctgcttg
//
5881 tttgtgcagc gactatgttg gtgttagggg tggtgtggag attgttaatc ttgtataaag
5941 caattcaata aattgtttca aggtttccaa aacaaaaaaa aa //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **PSSLPEGLPAPLEKRVKELAQHRGADSGAEQSGPLWGVCLSCVIPALQQ**
GCPAQACSETSPL
Frame 2: PLLSPKACQHPWRSALRSWLSIEAQTRELSSQVRSGVYAYLASFLPCSKDALLKR
ARKLHLCEQGGRLKEPLQKLKEAIGRAMPEQMAKYQDECQAHTQAKVAKMLEEEK
DKEQRDRICSDEEEDEEKGGRRIMGPRKKFQWNDEIRELLCQVVKIKLESQDLER
NNKAQAWEDCVKGFLDAEVKPLWPKGWMQARTLFKESRRGHGHLTSSLRPHSSN
Frame 3: LFSPRRPASTPGEAR

Breast Tumour cDNA Library – Round 9 – Plaque F2

GAATTCAAGCCGGAGGTGCAGGTCCTGGTGCTTGATGGTCGAGGCCATCTCCTGGGCCGCCTGGCGGCCATCG
TGGCTAAACAGGTACTGCTGGGCCGGAAGGTGGTGGTCGTACGCTGTGAAGGCATCAACATTTCTGGCAATTT
CTACAGAAACAAGTTGAAGTACCTGGCTTTCCTCCGCAAGCGGATGAACACCAACCCTTCCCGAGGCCCCTAC
CACTTCCGGGCCCCCAGCCGCATCTTCTGGCGGACCGTGCGAGGTATGCTGCCCCACAAAACCAAGCGAGGCC
AGGCCGCTCTGGACCGTCTCAAGGTGTTTGACGGCATCCCACCGCCCTACGACAAGAAAAAGCGGATGGTGGT
TCCTGCTGCCCTCAAGGTCGTGCGTCTGAAGCCTACAAGAAAGTTTGCCTATCTGGGGCGCCTGGCTCACGAG
GTTGGCTGGAAGTACCAGGCAGTGACAGCCACCCTGGAGGAGAAGAGGAAAGAGAAAGCCAAGATCCACTACC
GGAAGAAGAAACAGCTCATGAGGCTACGGAAACAGGCCGAGAAGAACGTGGAGAAGAAAATTGACAAATACAC
AGAGGTCCTCAAGACCCACGGACTCCTGGTCTGAGCCCAATAAAGACTGTTAATTCCTCCAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGGGGGGCGGCCCCCCCC
AAAAAAAATTTTAACCCCCCGGGGGGCCCAAAAGGGGGGGGGGGGGGAAAAAAAAA (up primer)

LOCUS NM_012423 1142 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens ribosomal protein L13a (RPL13A), mRNA.** ACCESSION
NM_012423 VERSION NM_012423.2 GI:14591905

1 ctttttccaag cggctgccga ag**atggcgga ggtgcaggtc ctggtgcttg atggtcgagg**
61 **ccatctcctg ggccgcctgg cggccatcgt ggctaaacag gtactgctgg gccggaaggt**
121 **ggtggtcgta cgctgtgaag gcatcaacat ttctggcaat ttctacagaa acaagttgaa**
181 **gtacctggct ttcctccgca agcggatgaa caccaaccct tcccgaggcc cctaccactt**
241 **ccgggccccc agccgcatct tctggcggac cgtgcgaggt atgctgcccc acaaaaccaa**
301 **gcgaggccag gccgctctgg accgtctcaa ggtgtttgac ggcatcccac cgccctacga**
361 **caagaaaaag cggatggtgg ttcctgctgc cctcaaggtc gtgcgtctga agcctacaag**
421 **aaagtttgcc tatctggggc gcctggctca cgaggttggc tggaagtacc aggcagtgac**
481 **agccaccctg gaggagaaga ggaaagagaa agccaagatc cactaccgga agaagaaaca**
541 **gctcatgagg ctacggaaac aggccgagaa gaacgtggag aagaaaattg acaaatacac**
601 **agaggtcctc aagacccacg gactcctggt ctga**gcccaa taaagactgt taattcctca
661 tgcgttgcct gcccttcctc cattgttgcc ctggaatgta cgggacccag gggcagcagc
721 agtccaggtg ccacaggcag ccctgggaca taggaagctg ggagcaagga aagggtctta
781 gtcactgcct cccgaagttg cttgaaagca ctcggagaat tgtgcaggtg tcatttatct
841 atgaccaata ggaagagcaa ccagttacta tgagtgaaag ggagccagaa gactgattgg
901 agggccctat cttgtgagtg gggcatctgt tggactttcc acctggtcat atactctgca
961 gctgttagaa tgtgcaagca cttggggaca gcatgagctt gctgttgtac acagggtatt
1021 tctagaagca gaaatagact gggaagatgc acaaccaagg ggttacaggc atcgcccatg
1081 ctcctcacct gtattttgta atcagaaata aattgctttt aaagaaaaaa aaaaaaaaaa
1141 aa //

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **RPGKVQVLVLDGRGHLLGRLAAIVAKQVLLGRKVVVVRCEGINISGNFYRNKL**
KYLAFLRKRMNTNPSRGPYHFRAPSRIFWRTVRGMLPHKTKRGQAALDRLKVF
DGIPPPYDKKKRMVVPAALKVVRLKPTRKFAYLGRLAHEVGWKYQAVTATLEE
KRKEKAKIHYRKKKQLMRLRKQAEKNVEKKIDKYTEVLKTHGLLV

Frame 2: GPGRCRSWCLMVEAISWAAWRPSWLNRYCWAGRWWSYAVKASTFLAISTETS
Frame 3: AREGAGPGA

Breast Tumour cDNA Library – Round 9 – Plaque H2

LOCUS NM_181786 2952 bp mRNA linear PRI 13-MAR-2011
DEFINITION **Homo sapiens HKR1, GLI-Kruppel zinc finger family member (HKR1), mRNA.**
ACCESSION NM_181786 VERSION NM_181786.2 GI:34330189

Gene in backwards

Breast Tumour cDNA Library – Round 9 – Plaque C3

TTCAAGTCCTGACATTCAATTCAAGAAGCTGGAAAGGTACAATAAAATACTCTCAGAAAAGCAAAGGAAATTA
CAAAAGAAAATATATAAATTAGGCCTGGCGCGGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCTGAGG
CAGGTGGATCACGAGATCAGGAGTTCAAGACCAGCCTGGCCAAGATGGTGAAACCCCATCTCAACTAAAAATA
CAAAAATTAGCTGGGCGTGGCGGCGTGCACCTGTAGTCCCAGCTTCTTGGGAGGCTGAGACAAGAGAATTGCT
TGAACCCGGGAGGTGGAGGTTGCAGTGAGCCCAGATCGCGCCACTGCACTCCAGCCTGGGTGACAGAGCGAGA
CTCTGTCTCAAAACAAAACAAAACAAAAAACGAAAATATATAAATTAATGATTTAGAAAAATAAAGTAGAACT
GATAAAATTAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTTTGGAGGGG
TTA (up primer)

About 50 hits, although max identity 91%, overlay only between bp 96-400 of query.

Frame 1: FKS
Frame 2: SSPDIQFKKLERYNKILSEKQRKLQKKIYKLGLARWLMPVIPALWEAEAGGS
RDQEFKTSLAKMVKPHLN
Frame 3: QVLTFNSRSWKGTIKYSQKSKGNYKRKYIN

Protein BLAST of frame 2 protein reveals about 50 similarities, but none more than 20 aa overlap.

Breast Tumour cDNA Library – Round 9 – Plaque D3

TAGGCGTCAGCCCTCTTCTCTCCCCGAAGGCCTGCCAGCACCCCTGGAGAAGCGCGTTAAGGAGCTGGCTCAG
CATAGAGGCGCAGACTCGGGAGCTGAGCAGTCAGGTCCGCTCTGGGGTGTATGCCTATCTTGCGTCATTCCTG
CCCTGCAGCAAGGATGCCCTGCTCAAGCGTGCTCGGAAACTTCACCTCTGTGAACAGGGGGGCCGTCTGAAGG
AGCCTCTCCAGAAGCTCAAGGAAGCCATTGGCAGGGCGATGCCAGAGCAGATGGCCAAGTACCAGGACGAATG
CCAGGCACACACGCAGGCAAAGGTTGCCAAGATGCTGGAAGAGGAGAAAGACAAGGAGCAGAGGGACCGGATT
TGTTCGGATGAGGAAGAAGATGAAGAAAAAGGGGGCAGGAGGATAATGGGACCTCGGAAGAAGTTCCAATGGA
ATGATGAGATCAGGGAGCTACTGTGCCAGGTGGTGAAGATCAAACTGGAGAGCCAGGACCTGGAGAGGAACAA
CAAAGCCCAGGCTTGGGAGGACTGTGTGAAGGGCTTTCTGGATGCGGAAGTCAAGCCCCTCTGGCCCAAAGGC
TGGATGCAGGCCAGAACTCTGTTTAAGGAGAGCAGACGAGGCCATGGGCACCTGACTTCAAGCTTGCGGCCGC
ACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGCTTGTGAAGGGGTTA (up primer)

LOCUS NM_001079514 5982 bp mRNA linear PRI 25-SEP-2011
DEFINITION **Homo sapiens ubinuclein 1 (UBN1), transcript variant 2, mRNA.**
ACCESSION NM_001079514 VERSION NM_001079514.1 GI:118572602

Phage cDNA insert starts with later part of gene, then shifts to previous part of gene - second part out of frame.

```
   1 gtcccgcgcc gcaagttctc ctggggcgac cggaggctgc tcccccgacc ccctggtgtc
  61 cccggagtgg ctgcgcggac gtcgagttgg catttcttcg cttcctcctg ggccggcgcg
 121 cggcgcggca caccaggctc ccctgggctc ggggacccgg ccatgggccg aggcgcgggc
 181 cgcccgcccg ctgggagcca cggcttagca gccgaccgct agctgcgccg ccgcccgggg
```

```
 241 accggcatga ggaccgccgc gggggacgt ctgcggcccg cgtcggcgct ggggacaaag
 301 aagccatgca gtgacacccg ctaagacttg ttggtagcca tgtcggagcc ccacagggtc
 361 cagttcacct ctctcccagg ttccctgaat cctgcgtttt tgaagaagtc ccggaaggag
//
1261 atggactcgc tgacggattt ggacttggag catctgctca gtgagtctcc agaaggaagt
1321 cccttccgag atatggatga tggaagtgat tcccttgggg tgggattgga ccaggaattc
1381 aggcagccct cttctctccc cgaaggcctg ccagcacccc tggagaagcg cgttaaggag
1441 ctggctcagg ctgccagagc tgctgagggg gagagcagac agaagttctt cacccaggat
1501 attaatggaa tcctattaga catagaggcg cagactcggg agctgagcag tcaggtccgc
1561 tctggggtgt atgcctatct tgcgtcattc ctgccctgca gcaaggatgc cctgctcaag
1621 cgtgctcgga aacttcacct ctatgaacag gggggccgtc tgaaggagcc tctccagaag
1681 ctcaaggaag ccattggcag ggcgatgcca gagcagatgg ccaagtacca ggacgaatgc
1741 caggcacaca cgcaggcaaa ggttgccaag atgctggaag aggagaaaga caaggagcag
1801 agggaccgga tttgttcgga tgaggaagaa gatgaagaaa aaggggcag gaggataatg
1861 ggacctcgga agaagttcca atggaatgat gagatcaggg agctactgtg ccaggtggtg
1921 aagatcaaac tggagagcca ggacctggag aggaacaaca aagcccaggc ttgggaggac
1981 tgtgtgaagg gctttctgga tgcggaagtc aagcccctct ggcccaaagg ctggatgcag
2041 gccagaactc tgtttaagga gagcagacga ggccatgggc acctgacttc aatcctggcc
2101 aagaagaaag ttatggcccc ttctaaaatc aaggtgaagg aatcgtctac gaagcctgat
2161 aaaaaggttt ctgtcccatc aggacagatt ggtggcccca ttgctttgcc ctcagatcac
//
3661 gcggtgccca cccatatccc gcagagtctg ccaggtgctt ctcagcttca cgggaaaggg
3721 cctgctgtac cacggaaatt gtgaccgctt cagaggcaag gcttgccact tgggtctggg
3781 tggaatcaga acgtgcaggt ctcccaggat gtacactcac tgcgcccttt ctgctgcttg
//
5881 tttgtgcagc gactatgttg gtgttagggg tggtgtggag attgttaatc ttgtataaag
5941 caattcaata aattgtttca aggtttccaa aacaaaaaaa aa //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **PSSLPEGLPAPLEKRVKELAQHRGADSGAEQSGPLWGVCLSCVIPALQQ**
GCPAQACSETSPL
Frame 2: PLLSPKACQHPWRSALRSWLSIEAQTRELSSQVRSGVYAYLASFLPCSKDALLKR
ARKLHLCEQGGRLKEPLQKLKEAIGRAMPEQMAKYQDECQAHTQAKVAKMLEEEK
DKEQRDRICSDEEEDEEKGGRRIMGPRKKFQWNDEIRELLCQVVKIKLESQDLER
NNKAQAWEDCVKGFLDAEVKPLWPKGWMQARTLFKESRRGHGHLTSSLRPHSSN
Frame 3: LFSPRRPASTPGEAR

Breast Tumour cDNA Library – Round 9 – Plaque E3

```
ATTTCAGTCTAAAGAAGCAGGTTTGGGAAGGAGGGAAGGAGGGAAGGGAGGAAGGAAAAGAAAGAAGCAGGTT
TGGGAAGAAAGATTCAGTTTTGGCTGGGCGTGGTGGCTCATGCCTGTAATCCCAGTGCTTTGGGAGGCCGAGG
CAGGTGGATCTCCTGAGGTCAAGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAA
TACAAAATAAATAAATAAATAACTAAAAAATAAGCTGGGCATGGTGGCAGGTGCCCGTAATCCCAGCTACTTG
GGAGGCTGAGGCAAGAGAATTGCTTGAAGCCAGGAGGTGGAGGTTGCAGTGAGCTGAGATTGCACCACTGCAC
TCCAGCCTGGGTGACAGAGTGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAAAAAATTCGGTTTCCTTTTA
CCCCTTGATTTTTGGGGGGGAAAAAAAAACCCATTACCCCTTCCCCCCCACCATTTGCAAACATTTTTCTAGG
GGGGGAAAGTATCCTGAGTGCATTCTGGGAAAATTCACAAAAAATTAAAAATGGGACAGGGTTTTAAAGGAGG
GGTAATTCAAAGAAGGGGAGGACTTTGTATTTGGTTAAACCCTTGCAAAATTGCTAGATGTCTCTATAGAAAC
CTTGCGCCCCCCCTCGAGAAATTATTAAACCCCTGGGGGCCTCTAAACGGTTTTGGGGGGGGTTACCC (up
primer)

About 50 hits, although max identity 91%, overlay only between bp 96-400
of query.

Frame 1: ISV
Frame 2: FQSKEAGLGRREGGKGGRKRKKQVWEERFSFGWAWWLMPVIPVLWEAEAGGSP
EVKSSRPAWPTW
```

Frame 3: FSLKKQVWEGGKEGREEGKERSRFGKKDSVLAGRGGSCL

Protein BLAST of frame 2 protein reveals about 70 similarities , but none more than 20 aa overlap.

Breast Tumour cDNA Library – Round 9 – Plaque F3

TTAAGGCTGGAGCAAAGAGACATGAAGAGAGGCCAGCGAGGTACGGGCAGAAGCTCAGCAGAACCAAGGCAGA
GACAAAAAAAAGACAAAGACAGATGGAAAGAGAGCTGGGGCTGGGCGTGGTGGCTCACGCCTGTAATCCCAGCA
CTTTGGGAGGCCGAGGCAGGTGAATCGCCTGAGGTCAGGAGTTCGAGACCAGCCTGGTCAACATGGTAAAACC
CGTCTCTACGAAAAATACAAAAATTAGCCGGGAGTAAGGGTGGGTGCCTTTAATCCCAGCTACTCAGGAGGCT
GAGTCCAGAGAATCGCTTGAACCCGGGAGGCAGAGGCGGGTGGATCATGAGGTCAAGAGATCAAGACCATTCT
TGCCAACATGGGGAAACCCCATCTCTACCAAAAATAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTT
GGGGCCTCTAAACGGGTCTTGAGGGGTTA (up primer)

About 50 hits, although max identity 90%, overlay only between bp 96-320 of query.

Frame 1: LRLEQRDMKRGQRGTGRSSAEPRQRQKKTKTDGKRAGAGRGGSRL
Frame 2: GWSKET-REASEVRAEAQQNQGRDKKRQRQMERELGLGVVAHACNPSTLGGRGR
Frame 3: KAGAKRHEERPARYGQKLSRTKAETKKDKDRWKESWGWAWWLTPVIPALWEAEA
GESPEVRSSRPAWSTW

Protein BLAST of frame 3 protein reveals about 70 similarities, but none more than 20 aa overlap.

Breast Tumour cDNA Library – Round 15 – Plaque A5

CAGTGGCAAGGTTTGATTGGGAAC**GAATTCT**CTTCCTCCTGTCCTCTGCTCTTGGGCCAGGTTCCCCCTCGTG
CATGAAACTAGAAAGCAGAAGGTAGGGAAGTTTATTGATGTGACCACACGGCTGGCCTTCTGATGAGAAGCAG
AGTGGAGACTGCATCTGGAGCACAAACTGGAGAGATCCAGAATACACAGGAAATTCTCTGCCTCCAAGGAGCT
CCCAATCTAGTAGGGAGACAGTGGAGTTAGCAGATATTCCAACATAATTTGAACAATAGTGACAAATTATTGA
GAACGCTCACAAGAACCCTATAAGATAGGTTATTGTGATGATTTTACAGAGGGAAGCTGAGAGACGGAGAGTT
TAAGGCATTTGCCTAGGTCACACAGTTGGGAAGTAAAGGGGTCGAGATTTGAACTTGGGAGTTTGTCTCCAGA
GTCTGTGCTTTGCTGGATTGGAAGAGTTAACCTGTGGTTTAGAGCAGTGCTTTTTTTTTTTTTTTTTGGAAA
TGGAGTTTTGCACTGTCCCCAGGGATGGAGGGCAGGGGGGGGATCTCAGTTCACTGCAACCCCTGCCTCCTGG
GTTCAAGTGATTCTCCTGCCTCACCCTCCCTAGTAGCTGGGGTTACAGGCACCAGCCCCCACACCCAGGTAAT
TTTTTTTTTTTTTTGTAGGACGGGGGTTTCATTATTTTGGGCAGGGTGGTCTAAAACTCCTGACCCCGGGAAC
CACCTCCCTTGGCCTCCCAAAGGGGTGAAATTACAGGGCTAAGCCCCTGCCCTTGCCAAAACATGGTTTTTAA
GCCTTCCGTTTTTGGGGAATTTTCTTGGGGGAGGTGGTTTAAAAAAATTCCTGGTTCCCGGGTCATTAAGCTA
GGGC (up primer)

Query coverage 33%, Max ident 80% - not a good match

LOCUS NM_001031701 7214 bp mRNA linear PRI 29-APR-2012
DEFINITION **Homo sapiens 5'-nucleotidase domain containing 3 (NT5DC3), mRNA.**
ACCESSION NM_001031701 VERSION NM_001031701.2 GI:219689102

CDS 42-1688
1 gcagagcggc gcggcggcgg tggcggcagc aggcagcagg c**atgaccatg gcagcggcgg**
61 **cggtggtggc acgcggggcc ggggcgaggg cagcgacagc ggcggctttg cggggtggct**
121 **gcgggaccgc ggctcggggg cggccgtgtg cgggccccgc ccggcccttg tgcactgcac**
//
1501 **cctacttcct aaggcgcctg tcgcgcttcg ctgacatcta catggcgtct ctgagctgcc**
1561 **tcctgaacta tgacgtcagc cacactttct acccccggag gactccactg cagcacgaac**
1621 **tgcccgcctg gtcagaaagg ccccccacct tcggaacccc tctcctgcag gaggcccagg**
1681 **ccaagtag**cc aagggcaaaa actagaaact gtaactgccc ctgattgggc aggcatgatg

1741 gggttgactt catgtgggtc tgagtgttgc ttttggaaaa gatacagata agccttttga
1801 tatttatttt gtaccttacg gagaataatt ctcccaatgg ttaagacagg agctagcagc
  //
4681 ggtttcacca tgttggccag gctggtctca aactcctagc ctcaagtgtt ctgcccacct
4741 cagccttcca aagtgctggg attctgggtg tgagccactg tcatgcctgg cctcgtcatg
4801 actcttttct ctttttttt ttttttttt ttgagattta gttttgctct ttttgcccag
4861 gccggagtgc aatggcacga tcttggctca ctccaacctc cgcctcctgg gttcaagtga
4921 ttctcctgcc tcagcctccc gagtagctgg gattacaggt gcccaccacc acatccagct
4981 aatttttttg tatttttagt agagacgggg tttcaccatg ttggccaggc tggtctcgaa
5041 ctcctgacct tgggtaatcc acctgccttg gcctcccaag gtgctgggat tacaggcatg
5101 agccaccgtg cccagcctat catgactcaa ctgggatctg gcacactgca ctgggtgttt
5161 tgttctctga tgtgcacatt gatgaatgca tattggtaca agggaaatga atctgttttc
  //
7141 attttaaatg ttgattttat gtgttatgtg actctgaact tattgccaaa taaaagtttg
7201 aattgtatca tctg

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: SLPPVLCSWARFPLVHETRKQKVGKFIDVTTRLAF
Frame 2: LFLLSSALGPGSPSCMKLESRR
Frame 3: SSSCPLLLGQVPPRA

BLASTp on frame 1 protein results in tens of matches, only for 5-10 aa per match

Breast Tumour cDNA Library – Round 15 – Plaque B5

GCGGAAATGCCCTGGCAC**GAATTCA**ACCAATCATGGATCAAATATATTCATGGGAAAAAAGGATGGCTAAGCC
TGTACTGAGGATGCACAGATTTTTTTTCCCTGTCATTCCCTAAGCAATACAATTCCCTAACAATAACAACTAT
TTACATAGCATTTACGTTGTAGTCGGTATTATAAGCAATCTAGAGATTATTTAAAGCATACAGAAGGATGTGC
CTAGGATATGTGCAAATACTGTAGGGACTTGAACATCCATGGATTTAGGTATCTATCAGGGGTGGGGAGCACT
CTAGTACAAGCAGACAGTCTCCTCTGGTGTGTAACTTACCACAGGTAGGGTGAGCCATGTTGCCACAAGGCGG
TTGCTGATCCAGCGATACCAAGATGGGTTTACAAACATCAAAGGTAAAAAGGGACTCAGCATGAAAATGCTTC
CAAAAAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTTTGAGGGGTTA
(up primer)

LOCUS NM_182551 5077 bp mRNA linear PRI 21-APR-2012
DEFINITION **Homo sapiens lysocardiolipin acyltransferase 1 (LCLAT1), transcript variant 1, mRNA.**
ACCESSION NM_182551 XM_171014 VERSION NM_182551.3 GI:50659060

Gene in backwards

Breast Tumour cDNA Library – Round 15 – Plaque D4

TAGGGGAATTTTTAGGAC**GAATTCA**AGCGCCTCTTCCTCTCCTTCGCCTTCTTGCCCTACATCAGCTTTGGCA
AGTTCGACCTGTACCGGAAACGCTGCCAGATCATCATCTTTCAGGTGGTCTTCCTGGGCCTCCTGGCTGGCCT
GGTGGTCCTCTTCTACGTCTATCCTGTCCGCTGTGAGTGGTGTGAGTTCCTCACCTGCATCCCCTTCACTGAC
AAGTTCTGTGAGAAGTACGAACTGGACGCTCAGCTCCACTGAGCTGGCTGCGGGCTCCAGCGGCCGTGTGCTC
CAGCAGGCCAGAGCCAGACACGACCTCCCTGGAAAGCATGAACCCAGGAGGCAGAGCTTGCAGTGAGCCAAAT
AGGGCCACTGCACTCCAGCCTGGGAGACAGAGCTAGACTCCACTCCATCTCAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAGTTGGGGCCCCCCCCCAAAAAATTATTAACCCCCTGGGGGCCTTAAAAGGGGTTTGGGGGGT
TAAAAGG (up primer)

LOCUS NM_022450 3091 bp mRNA linear PRI 26-MAR-2012
DEFINITION **Homo sapiens rhomboid 5 homolog 1 (Drosophila) (RHBDF1), mRNA.**
ACCESSION NM_022450 VERSION NM_022450.3 GI:190341096

CDS 149-2716
1 gcattcccgg gcgggccgga ccggcgggcg ggcggggact cggcgcgggc gccctcccgg
61 ccagcggcgg cagcccctcc tccccggcgc cctcaggacc ccccagagac ccccggcggc
121 ggcagcctgc cttgctctgc caggaacc**at gagtgaggcc cgcagggaca gcacgagcag**
181 **cctgcagcgc aagaagccac cctggctaaa gctggacatt ccctctgcgg tgcccctgac**
241 **ggcagaagag cccagcttcc tgcagcccct gaggcgacag gctttcctga ggagtgtgag**
//
2341 **cgtggagctc ttccagagct ggcagatcct ggcgcggccc tggcgtgcct tcttcaagct**
2401 **gctggctgtg gtgctcttcc tcttcacctt tgggctgctg ccgtggattg acaactttgc**
2461 **ccacatctca gggttcatca gtggcctctt cctctccttc gccttcttgc cctacatcag**
2521 **ctttggcaag ttcgacctgt accggaaacg ctgccagatc atcatctttc aggtggtctt**
2581 **cctgggcctc ctggctggcc tggtggtcct cttctacgtc tatcctgtcc gctgtgagtg**
2641 **gtgtgagttc ctcacctgca tccccttcac tgacaagttc tgtgagaagt acgaactgga**
2701 **cgctcagctc cactga**gctg gctgcgggct ccagcggccg tgtgctccag caggccagag
2761 ccagacacga cctccctgag cctcacaggc ttacaggagt cacctgctcc atgtggggac
2821 tggcctgttt cctgaacaca gacctctttc ttgtgccttg ttcacttctg ttgaacccct
2881 cgtactgccg ggcatttatt atactacttc ctgtcataac cttctaactt gtttcttgac
2941 gaccacctca tgtggccaat aaatggactg ggagcgtttt agctgccatt aacttgaaaa
3001 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
3061 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa a

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: SSASSSPSPSCPTSALASSTCTGNAARSSSFRWSSWASWLAWWSSSTSILSAVSGVS
SSPASPSLTSSVRSTNWTLSSTELAAGSSGRVLQQARARHDLPGKHEPRRQSLQ
Frame 2: QAPLPLLRLLALHQLWQVRPVPETLPDHHLSGGLPGPPGWPGGPLLRLSCPL
Frame 3: KRLFLSFAFLPYISFGKFDLYRKRCQIIIFQVVFLGLLAGLVVLFYVYPVRCEWCEF
LTCIPFTDKFCEKYELDAQLH

BLASTp of frame 1:
Homology with best matched protein is exclusively the serines present!

BLASTp of frame 2:
Homology not more than 12 aa. E value always above 2e-3

Breast Tumour cDNA Library – Round 15 – Plaque G5

TATAAGACACCTCGGGCTC**GAATTCT**TTAGTTTATCGGATTATTGAAGGAAAGCTGCCAGCAATGCCAAGAGA
TTACAGCCCAGAGCTGGCAGAACTGATAAGAACAATGCTGAGCAAAAGGCCTGAAGAAAGGCCGTCTGTGAGG
AGCATCCTGAGGCAGCCTTATATAAAGCGGCAAATCTCCTTCTTTTTGGAGGCCACAAAGATAAAGCTTGCGG
CCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTCACTAGTTACTCGAGT
GCGCCAGCTTTATCTTTGTGGCCTCCAAAAGAAGGAGATTTGCCCTTTATATAAGGCGCCTCGGATCTCTCCA
GACGGCTTTCTTCGGCTTTTGCTCACATGTTCTTATCTTTCTGCGCTCTGGGCTGTAATCTCTTGGATTCTGG
CGTTTTCTTCATAATCCATAAACTAAGACTTCGATCCCCGCATCCCCTGACTGGAATCCACGCTCCAATCTCT
(up primer)

LOCUS NM_003157 3745 bp mRNA linear PRI 01-APR-2012
DEFINITION **Homo sapiens NIMA (never in mitosis gene a)-related kinase 4 (NEK4), transcript variant 1, mRNA.**
ACCESSION NM_003157 VERSION NM_003157.4 GI:302058310

CDS 204-2729

```
   1 agcatgcgca gaactgctcc cggcccggat cgctatggca gcggcgtcgt cgcgggccgg
  61 gccccagcaa tcccgcccgg gcccggctgc ctcaacagcc gccccactg ccccctctcg
 121 ggcatgaacc gagcttcttg ttgccgcccg ctgccctacc cgccgctgcc gccgcatccc
 181 gactctgggc cagcgctggg aacatgcccc tggccgccta ctgctacctg cgggtcgtgg
 241 gcaaggggag ctatggagag gtgacgcttg tgaagcaccg gcgggacggc aagcagtatg
  //
 721 tgagccctga attgttctca aacaaaccct acaactataa gtctgatgtt tgggctctag
 781 gatgctgtgt ctatgaaatg gccaccttga agcatgcttt caatgcaaaa gatatgaatt
 841 ctttagttta tcggattatt gaaggaaagc tgccaccaat gccaagagat tacagcccag
 901 agctggcaga actgataaga acaatgctga gcaaaaggcc tgaagaaagg ccgtctgtga
 961 ggagcatcct gaggcagcct tatataaagc ggcaaatctc cttctttttg gaggccacaa
1021 agataaaaac ctccaaaaat aacattaaaa atggtgactc tcaatccaag ccttttgcta
1081 cagtggtttc tggagaggca gaatcaaatc atgaagtaat ccacccccaa ccactctctt
1141 ctgagggctc ccagacatat ataatgggtg aaggcaaatg tttgtcccag gagaaaccca
  //
2581 agcttttaga gcaggtgtat gatcttttgg aggaggagga tgaatttgat agagaggtac
2641 gtttgcggga gcacatgggt gaaaagtata caacttacag tgtgaaagct cgccagttga
2701 aatttttga agaaaacatg aatttttgag catttgtcct aatctgctgc cagaattaaa
2761 gacctatttt tagaggattt tggcttaaaa agcaagggca aacagtcatt tggaagccac
  //
3661 ttgaatagta cattttcact tctgtattcg agggactcta aaaataaata ttgctccaga
3721 aatgttaaaa aaaaaaaaaa aaaaa
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: EFFSLSDY
Frame 2: NSLVYRIIEGKLPAMPRDYSPELAELIRTMLSKRPEERPSVRSILRQPYIKRQISFF
LEATKIKLAAALE
Frame 3: IL

Breast Tumour cDNA Library – Round 15 – Plaque H5

ACCCAGGGAAATTTCCGGGACC**GAATTCT**CAAGCCTCCCTTATCTCAGGATGCTGCACTCCCACCGCCTCCTA
CAGTTATTCTTGAAAACTACAGTGAGAAATCGGGGCAAACTGGGTCTGCCCAAGGATCCCCAGAGAACCGCCC
AGCCATATTCAATGCTTTCTTATTTTTTTTTTTCCAAATCCGTAAGTTACCTGGTTTAATATCAACATTCTAA
TATTTTTAAAGGCTCAGTAGGGGGTTTGCCAAATTGTATGTTTATTCCCTTTTTTATTCTGTATGTTTAATAT
ATCCATAGTAAAACAATTACTTAAAAACAACTGAGGACTTTATGCAACGTGAGTGACAATGAATCTGAGAAGC
AAAACACACATGCTGCCCCTTCCGTAGTCTCTTTACAGTCCAGGCCTCCATGTACTCAGCAACTGCAGAATGA
CCCTCCAGCCCCAAACAACCATATGGCCATACGTGGAGCCCCTCCATAGGGTTGTAGCCTGGGTTGGAGGGAC
CTGCAGTACAGTATCTGGAACTCACAGCAAGGGGCCAGGCCCAAAGCAGGAGAAAATTTGAGAAACAAGATTG
CAAGAAACTGTCCAGGGAAACATGGCAAAACCTCCTCTCTACTAAAAATACAAAAAATGAGCTGGGCATGGTG
ATGTCTGCCTGTAGTCCCAGCTACTTGGGAGGCTGAGTGGGGAGAATCAGCTGAGACTCAGAGGTCGAGACTG
CAGTGAGCTGGGATCATGCCACTGCACTTCATCCTGGGCAACCAAAGTGAGACCCTGTCTCAAAAAAAAAAAA
AAGCTTGCGGCCGCACTCGATAACTAGTTAACCCCTTGGGGGCCTCTAAACGGGTCTTTGAGGGGTTAA (up primer)

Max query coverage <23%, Max Ident <82%

LOCUS NM_001193421 11671 bp mRNA linear PRI 01-APR-2012
DEFINITION **Homo sapiens teashirt zinc finger homeobox 2 (TSHZ2), transcript variant 2, mRNA.**
ACCESSION NM_001193421 XM_002343726 XM_002345419 XM_002348048 VERSION NM_001193421.1 GI:301171535

CDS 364-3459, but query coverage only approx 200bp from 10422-10626

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **EFSSLPYLRMLHSHRLLQLFLKTTVRNRGKLGLPKDPQRTAQPYSMLSYFFFSKSVSYLV**
Frame 2: NSQASLISGCCTPTASYSYS
Frame 3: ILKPPLSQDAALPPPPTVILENYSEKSGQTGSAQGSPENRPAIFNAFLFFFFQIRKLPGL ISTF

BLASTp of frame 1:

results in about 40-50 "matches" with:
query coverage <61%, Max Ident up to 100%, but actually only max. 9 out of 15 aa match up.

8.3.3.2.7 Pseudomonas stutzeri

***Pseudomonas stutzeri* gDNA Library – Round 9 – Plaque B7**

GGCTCCGGGGTAACGCATCCAGGCTCACGCCGGGCGGCGGATTGTCACGCAGCTGGTCGCCCAGCTGACGGCC
GATGCCATAGCTTACGCGCGTTTCGTCGGTGGACAGATTAAGGTCGGTCATGCCTGGGCTCCGCTCTGGAAAA
AGGTCGCACAGCTTAGGGGTGCGACCGTTTCGTCGCAAGCCGCGGACGAAACCGCAACAGCCCCTGGCACAGC
GTACCCGGCGAACCGCCTCTCAATGTTTGGTGAGTTTGTCCAGATAGCCCATGGCAAAGGCCGAGAAAACGAA
GGTCAGGTGAATGATCCGAATTCTCCTGCAGGGATATCCCGGGAGCTCGTCGACAAGCTTGCGGCCGCACTCG
AGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTTTGAGGGGTTAAC (forward primer)

LOCUS NC_009434 4567418 bp DNA circular BCT 16-NOV-2011
DEFINITION **Pseudomonas stutzeri A1501 chromosome, complete genome.**
ACCESSION NC_009434 VERSION NC_009434.1 GI:146280397

Features in this part of subject sequence:
peptidyl-prolyl cis-trans isomerase, FKBP-type
Score = 165 bits (182), Expect = 6e-41 Identities = 122/140 (87%), Gaps = 2/140 (1%) Strand=Plus/Minus

```
Query  14       CGCATCCAGGCTCACGCCGGGCGGCGGATTGTCACGCAGCTGGTCGCCCAGCTGACGGCC  73
                |||  ||  |||||  ||  ||  ||||||||  ||||||||||||||  ||  ||||||||||||||
Sbjct  1059230  CGCGTCGAGGCTGACACCCGGCGGCGGGTTGTCACGCAGCTGATCACCCAGCTGACGGCC  1059171

Query  74       GATGCCATAGCTTACGCGCGTTTCGTCGGTGGACAGATTAAGGTCGGTCATGCCTGGGCT  133
                ||||||  |||||  |||||  ||||||||||||||  |||||  ||  ||||||||||||  |||||
Sbjct  1059170  GATGCCGTAGCTGACGCGGGTTTCGTCGGTGGAGAGATTGAGTTCGGTCATGCC-GGGCT  1059112

Query  134      CCGCTCTGG-AAAAAGGTCG  152
                |||  ||  ||  ||||||||||
Sbjct  1059111  CCGGTCAGGTAAAAAGGTCG  1059092
```

Features in this part of subject sequence:
hypothetical protein
Score = 73.4 bits (80), Expect = 3e-13 Identities = 58/70 (83%), Gaps = 0/70 (0%) Strand=Plus/Minus

```
Query  239      TCAATGTTTGGTGAGTTTGTCCAGATAGCCCATGGCAAAGGCCGAGAAAACGAAGGTCAG  298
```

```
               ||| || |||||||  |||||||  || |||||||||||||||| |  |||||||| ||
Sbjct  1056959  TCAGTGCTTGGTGATCTTGTCCATGTAACCCATGGCAAAGGCCGACACCACGAAGGTGAG  1056900

Query  299      GTGAATGATC  308
                ||| ||||||
Sbjct  1056899  GTGGATGATC  1056890
```

Peptide encoded by Phage DNA:

Frame 1: GSGVTHPGSRRAADCHAAGRPADGRCHSLRAFRRWTD
Frame 2: APG
Frame 3: LRGNASRLTPGGGLSRSWSPS

***Pseudomonas stutzeri* gDNA Library – Round 9 – Plaque G9**

GGATCCGGGTATGCGCATACCGAGGCATCTGGTGACGGCCCCGGCGTGGCCGGCTCATCCAGCCTGGCTGGCG
GGCTTTCTCAAGCTGCTGGGCACCGAGATAAAACTGTAGGATGGCTGTCCTTTGCCGGCCCGGTGCCGGCAAG
CCCACCATTGTTAAGGAGCGCGATCACCCGGTCGGCCACGGTCCGGGCGAAACCCATTTCATGGGTCACGCAG
AGCATGGTCATGCCGCTTTCGGCGAGGCCGATCGCCCAGGTCGGCAGCCCGATGGACATCTACGAGACCCCGG
CTAGCCGGCTGGTCTGCGAGTTCATCGGCAACGTCAACCTGTTCGACGGCGAACTGATCGATTTCGGCTGCGG
CTCGGGCATTCTCGCCATCGCGGCGTTGCTGCTCGGCGCCGAACAGGCCATCGGCACCGATATCGATCACCTG
CTGGCGGTAGATGCCGAGGTTCTGCCGCTCCTTGTTGGGCCCCTTGGTGATGGTCAGGCCCCAGGTGATCAAG
GCGCACGGCGACCTGCACGATTTCATGCAATGGAAAGGGCCGATCCGAATTCTCCTGCAGGGATATCCCGGGA
GCTCGTCGACAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTTTGGAAGG
GGTTAATAC (forward primer)

Features in this part of subject sequence:
putrescine ABC transporter ATP-binding protein
Score = 152 bits (82), Expect = 7e-37 Identities = 94/100 (94%), Gaps = 0/100 (0%) Strand=Plus/Plus

```
Query  249    GATCGCCCAGGTCGGCAGCCCGATGGACATCTACGAGACCCCGGCTAGCCGGCTGGTCTG  308
              ||||||||||||||||||||||||||||||||||||||| ||||| || || ||||||||
Sbjct  88995  GATCGCCCAGGTCGGCAGCCCGATGGACATCTACGAGACGCCGGCCAGTCGCCTGGTCTG  89054

Query  309    CGAGTTCATCGGCAACGTCAACCTGTTCGACGGCGAACTG  348
              |||||||||||||||||||||||||||| ||||| |||
Sbjct  89055  CGAGTTCATCGGCAACGTCAACCTGTTCGAGGGCGAGCTG  89094
```

Putrescin ABC transporter ATP-binding protein:

```
   1 atggccgttg cctccagcgc ctacaagaag gccctcagtg gcgagagcaa gaacaagcag
  61 gtgctgctga agatcgatcg cgtcaccaag aagtttgacg aaacggtggc ggtcgatgac
 121 gtctcgctga gcattcatca gggcgagatc ttcgcgctgc tgggcggttc cggctcgggc
 181 aagtcgaccc tgctgcgcat gctggcgggc ttcgagcgac ccaccgaagg gcgcatcttc
 241 ctcgatggcc aggacatcac cgacatgccg ccgtacgagc ggccgatcaa catgatgttc
 301 cagtcctatg ccctgttccc gcacatgacc gttgagcaga acatcgcctt cggtctcaag
 361 caggacggtc tgcctaagac cgagatcgag gagcgggtca aggaaatgct cggtctggtg
 421 cagatgaccc agtacgccaa gcgcaaacca catcagctct ctggcggcca gcgccagcgc
 481 gtcgccctgg cgcgctcgct ggccaagcgc ccgaaactgc tgctgctcga cgagccgatg
 541 ggcgcgttgg acaagaagct gcgctcgcag atgcaactgg aactggtaca gatcatcgag
 601 cgcgtcggcg tgacctgcgt gatggtgacc cacgaccagg aagaggccat gaccatggcc
 661 gagcgcatcg ccatcatgca tcagggctgg atcgcccagg tcggcagccc gatggacatc
 721 tacgagacgc cggccagtcg cctggtctgc gagttcatcg gcaacgtcaa cctgttcgag
 781 ggcgagctgg tggaagacat gggcgaccac gcgatcatcg cctcgcccgg gctggacaat
 841 ccgatctacg tgggccacgg catcagcacc cgtgcgcagg acaagcacat cacctacgcc
 901 ctgcgcccgg aaaagctgct gatcggcacc gagctgccgg aactggagcg tccgggctac
 961 aactgggcca agggcgtcgt gcacgacatc gcctacctgg gcggtcattc ggtgtactac
1021 atcaagctac cgtccggtgg tgtgctgcag gccttcatgg ccaacgccga acgccatgtg
1081 aaactgccga cctgggagga agaggtgtac gtctactggt gggatgacag cggcgtggtg
1141 ctgcaggcat ga
//
```

```
Features in this part of subject sequence:
3-octaprenyl-4-hydroxybenzoate carboxy-lyase
Score =  134 bits (72),  Expect = 3e-31  Identities = 83/88 (94%), Gaps =
1/88 (1%)  Strand=Plus/Minus

Query  429     CGATCACCTGCTGGCGGTAGATGCCGAGGTTCTGCCGCTCCTTGTTGGGCCCCTTGGTGA  488
               ||||||||||||| |||||||||||||||||||||| |||||||| ||||||||||||
Sbjct  632123  CGATCACCTGCTGACGGTAGATGCCGAGGTTCTGCCGTTCCTTGTTCGGCCCCTTGGTGA  632064

Query  489     TGGTCAGGCCCCAGGTGATCAA-GGCGC  515
               ||||||||||||||||||||  |||||
Sbjct  632063  TGGTCAGGCCCCAGGTGATCAGCGGCGC  632036


Features in this part of subject sequence:
ThiJ/PfpI family protein   Score =  134 bits (72),  Expect = 3e-31
Identities = 96/107 (90%), Gaps = 4/107 (4%)  Strand=Plus/Plus

Query  13       GCGCATACCG-AGGC-ATCTGGTGACG-GCCCCGGCGTGGCCGGCTCATCCAGCCTGGCT  69
                |||||  |||| |||| ||||||| | | |||||||| |||||||| ||||| ||||||||
Sbjct  1115019  GCGCACACCGAAGGCAATCTGGTCA-GCGCCCCGGCTTGGCCGGCGCATCCCGCCTGGCT  1115077

Query  70       GGCGGGCTTTCTCAAGCTGCTGGGCACCGAGATAAAACTGTAGGATG  116
                |||||||||||||||||||| |||||||||||||| ||||||||||
Sbjct  1115078  GGCGGGCTTTCTCAAGCTGCTCGGCACCGAGATAAAGCTGTAGGATG  1115124


Peptide encoded by Phage DNA - BLASTp search:

Frame 1: GSGYAHTEASGDGPGVAGSSSLAGGLSQAAGHRDKTVGWLSFAGPVPASPPLLR
         SAITRSATVRAKPISWVTQSMVMPLSARPIAQVGSPMDIYETPASRLVCEFIGN
         VNLFDGELIDFGCGSGILAIAALLLGAEQAIGTDIDHLLAVDAEVLPLLVGPLG
         DGQAPGDQGARRPARFHAMERADPNSPAGISRELVDKLAAALE

      LOCUS    YP_004712448   383 aa     linear   BCT 29-NOV-2011
      DEFINITION  putrescine ABC transporter ATP-binding protein
      [Pseudomonas stutzeri ATCC 17588 = LMG 11199].
      ACCESSION   YP_004712448 VERSION   YP_004712448.1   GI:339492155

      Score =  109 bits (250),  Expect = 3e-22  Identities = 32/35
      (91%), Positives = 35/35 (100%), Gaps = 0/35 (0%)

      Query  84   IAQVGSPMDIYETPASRLVCEFIGNVNLFDGELID  118
                  IAQVGSPMDIYETPASRLVCEFIGNVNLF+GEL++
      Sbjct  231  IAQVGSPMDIYETPASRLVCEFIGNVNLFEGELVE  265


Frame 2: DPGMRIPRHLVTAPAWPAHPAWLAGFLKLLGTEIKL

      LOCUS    YP_004713419    193 aa   linear   BCT   29-NOV-2011
      DEFINITION  ThiJ/PfpI family protein [Pseudomonas stutzeri  ATCC
      17588 = LMG 11199].
      ACCESSION   YP_004713419 VERSION     YP_004713419.1   GI:339493126

      Score = 86.3 bits (196), Expect = 7e-23  Identities = 26/27
      (96%), Positives = 26/27 (96%), Gaps = 0/27 (0%)

      Query  10   LVTAPAWPAHPAWLAGFLKLLGTEIKL  36
                  LV APAWPAHPAWLAGFLKLLGTEIKL
      Sbjct  167  LVSAPAWPAHPAWLAGFLKLLGTEIKL  193


Frame 3: IRVCAYRGIW
```

[more than 50 matches, lowest E value = 0.34]

Pseudomonas stutzeri **gDNA library – Round 15 – Plaque C7**

CCTACGAATCCTGGGAACGCATCAGGCTCACGCCGGGCGGCGGATTGTCACGCAGCTGGTCGCCCAGCTGACG
GCCGATGCCATAGCTTACGCGCGTTTCGTCGGTGGACAGATTAAGGTCGGTCATGCCTGGGCTCCGCTCTGGA
AAAAGGTCGCACAGCTTAGGGGTGCGACCGTTTCGTCGCAAGCCGCGGACGAAACCGCAACAGCCCCTGGCAC
AGCGTACCCGGCGAACCGCCTCTCAATGTTTGGTGAGTTTGTCCAGATAGCCCATGGCAAAGGCCGAGAAAAC
GAAGGTCAGGTGAATGATCCGAATTCTCCTGCAGGGATATCCCGGGAGCTCGTCGACAAGCTTGCGGCCGCAC
TCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGT (up primer)

TATGTCTTTGCAAGCTTAGAAACTAATTAGCGTCCTTTATCTTTTTTTTTTCTTTCACAGCCTCCTACAGTTA
TTCTTGATAGAAAAAGAGAAAGGGGGCAAACTGGGTATGGCCAAGGATCCTCAGAAAACCGCCCATGATTATC
CAATGCTTTCTTCCTTTTTTTTTTCGAAATCCGGCAGATACCTGGTGCCCATCTAAATTTTAATATTTTTAAA
GGCTCAGAGGGGGGCCAGCCAAATCGGGACAGATTCCCTTTTTTATTCTGTATGATTAATATATCCATAGGGA
AACAATCCCTTCAAAGAACTCAGGACTTTATCCGCGTGACCTCCATGAATGCGACAACCCAAAGGACGGGCTC
CCCCTTCCCTATTCTCTTTTTTGGGAGGCCTCCATGTACTTGGCATCGGTAGAAACACCCTCCCTTGCCATGC
AAACAAACGGCCAAAGGGGGAGCCCCTCCGTTTGGCTGTAGCCAGGGATGGAGGGACCCGCTGCACAGAATCC
TCAAC (down primer)

DEFINITION **Pseudomonas stutzeri A1501 chromosome, complete genome.**
ACCESSION   NC_009434 VERSION   NC_009434.1 GI:146280397

Features in this part of subject sequence:
peptidyl-prolyl cis-trans isomerase, FKBP-type
Score = 161 bits (178), Expect = 2e-38 Identities = 117/133 (88%), Gaps = 2/133 (2%) Strand=Plus/Minus

```
Query  24       AGGCTCACGCCGGGCGGCGGATTGTCACGCAGCTGGTCGCCCAGCTGACGGCCGATGCCA  83
                |||||  ||  ||  ||||||||  ||||||||||||||  ||  ||||||||||||||||||||
Sbjct  1059223  AGGCTGACACCCGGCGGCGGGTTGTCACGCAGCTGATCACCCAGCTGACGGCCGATGCCG  1059164
Query  84       TAGCTTACGCGCGTTTCGTCGGTGGACAGATTAAGGTCGGTCATGCCTGGGCTCCGCTCT  143
                |||||  |||||  ||||||||||||||  |||||  ||  ||||||||||  ||||||||  ||
Sbjct  1059163  TAGCTGACGCGGGTTTCGTCGGTGGAGAGATTGAGTTCGGTCATGCC-GGGCTCCGGTCA  1059105
Query  144      GG-AAAAAGGTCG  155
                ||  ||||||||||
Sbjct  1059104  GGTAAAAAGGTCG  1059092
```

Features in this part of subject sequence:
hypothetical protein
Score = 73.4 bits (80), Expect = 6e-12 Identities = 58/70 (83%), Gaps = 0/70 (0%) Strand=Plus/Minus

```
Query  242      TCAATGTTTGGTGAGTTTGTCCAGATAGCCCATGGCAAAGGCCGAGAAAACGAAGGTCAG  301
                |||  ||  |||||||  |||||||  ||  |||||||||||||||  |  ||||||||  ||
Sbjct  1056959  TCAGTGCTTGGTGATCTTGTCCATGTAACCCATGGCAAAGGCCGACACCACGAAGGTGAG  1056900
Query  302      GTGAATGATC  311
                |||  ||||||
Sbjct  1056899  GTGGATGATC  1056890
```

Pseudomonas stutzeri **gDNA library – Round 15 – Plaque E8**

TTTTAACGAAATCCGCGTACGAGCAGCCGCGCTGCAGCGGGCCTGAACCTGACCGAAGTACGGACATCAGCCA
CTTACTCTATGGAAGGTCGTTCACATGGCAGCGCCAATCAGCAGTCCACGCTGGTACCTGATCCGAATTCTCC
TGCAGGGATATCCCGGGAGCTCGTCGACAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCT
CTAAACGGGTCTTGAGGGGTTAACTACTACTCGGAAGGTCGTTGACAGGGCAGAGCTTCGGGATATCGCTGGG
GATGTTCGATCCGCTGAGGGGGACCGGTGAGCGCGTCTACCGGTGGCGGACCCACTTAGAAAATAGGTATCAC
TTGGGGTTCTAAACGGGTTTTGAGGGGTTAGCGCGGGCTGCTCCTTCCGAGCATCCACCTGACGGAATACACG
CCCACA (up primer)

GGCAAATTTTTAACGTGTTACTCGAGTGCGGCCGCAAGCTTGTCGACGAGCTCCCGGGATATCCCTGCAGGAG
AATTCGGAGAGGTACCAGCGTGGACTGCTGATTGGCGCTGCCATGTGAACGACCTTCCATAGAGTAAGTGGCT
GATGTCCGTACTTCGGTCAGGTTTCAGGCCCGCTGCAGCGCGGGCTGCTCGATCCCCGAGCATCACACCTGAC

TGGAATACGACAGCTCCACTTACTACTCGGATGGGCTTCACTTGGGAGAGCTTCGGGACTCCTGGGTACTAAT
CCGATTTCCTG (down primer)

LOCUS JN874651 4307 bp DNA circular SYN 04-JAN-2012
DEFINITION **T7 Flexi Vector pFC30K His6HaloTag, complete sequence.**
ACCESSION JN874651 VERSION JN874651.1 GI:366091003

Pseudomonas stutzeri **gDNA library – Round 15 – Plaque F7**

TATAAGAGCCTCGCGTACCCGCATCCAGGCTCCGCCGGGCGGCGGATTGTCACGCAGCTGGTCGCCCAGCTGA
CGGCCGATGCCATAGCTTACGCGCGTTTCGTCGGTGGACAGATTAAGGTCGGTCATGCCTGGGCTCCGCTCTG
GAAAAAGGTCGCACAGCTTAGGGGTGCGACCGTTTCGTCGCAAGCCGCGGACGAAACCGCAACAGCCCCTGGC
ACAGCGTACCCGGCGAACCGCCTCTCAATGTTTGGTGAGTTTGTCCAGATAGCCCATGGCAAAGGCCGAGAAA
ACGAAGGTCAGGTGAATGATCCGAATTCTCCTGCAGGGATATCCCGGGAGCTCGTCGACAAGCTTGCGGCCGC
ACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGT (up primer)

TTCCTTTTTTTAACATGTTCTGTTTCTCGAGTGCGGCCGCAGCTTGTCGACAGCTCCCGGGATATCCCTGCAG
GAGAATTCGGATCATTCACCTGACCTTCGTTTTCTCGGCCTTTGCCATGGGCTATCTGGACAAACTCACCAAA
CATTGAGAGGCGGTTCGCCGGGTACGCTGTGCCAGGGGCTGTTGCGGTTTCGTCCGCGGCTTGCGACGAAACG
GTCGCACCCCTAAGCTGTGCGACCTTTTTCCAGAGCGGAGCCCAGGCATGACCGACCTTAATCTGTCCACCGA
CGAAACGCGCGTAAGCTATGGCATCGGCCGTCAGCTGGGCGACCAGCTGCGTGACAATCCGCCGCCCGGCGTG
AGCCTGGATGCGGTGATCCCCGAGCATCACACCTGACTGGAATACGACAGCTCCA (down primer)

Identical hit to plaque C7

Pseudomonas stutzeri **gDNA library – Round 15 – Plaque H7**

GCTCTGAGCCCCTTCCAAGGGATTCGCCATCAGGCTCACGCCGGGCGGCGGATTGTCACGCAGCTGGTCGCCC
AGCTGACGGCCGTTGCCATAGCTTACGCGCGTTTCGTCGGTGGACAGATTAAGGTCGGTCATGCCTGGGCTCC
GCTCTGGAAAAAGGTCGCACAGCTTAGGGGTGCGACCGTTTCGTCGCAAGCCGCGGACGAAACCGCAACAGCC
CCTGGCACAGCGTACCCGGCGAACCGCCTCTCAATGTTTGGTGAGTTTGTCCAGATAGCCCATGGCAAAGGCC
GAGAAAACGAAGGTCAGGTGAATGATCCGAATTCTCCTGCAGGGATATCCCGGGAGCTCGTCGACAAGCTTGC
GGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTAA (up primer)

GACACCAAGGTTTTACGTGATTTTCTCTCAGTGCTGCCGCAAGCTTGTCGACTAGCTCCCGGGATATCCCTGC
AGGAAAATTCGAATCATTCACCTGACCTTCGTTTTCTCGGCCTTTGCCATGGGTCTATCTGGACAAACTCTCC
AAACATTGAAAGGCGGTTCGCCGGGTACGTTGTGCCAGGGGCTGTTGCGGCTTCGACCACGGCTTGCGACGAA
ACGGTCGCACCCCTAAGGTGAACGACCTTTTTCCAGAGCGGAGCCCAGGCATGACCGACCTTAATCTGTCCAC
CGACGAAACGCGCGTAAGCTATGGCATCGGTCGTCTGCTGGGCTATCAGCTGCGTGACAATCCGCCGTCCGGG
GTGAGCCTGGATGCCGAGATCCCCCATCGTCACACCTGACTGGAATACGACAGCTCCAAGTT (down
primer)

Identical hit to plaque C7

8.3.3.3 Daptomycin

8.3.3.3.1 Human colon

Colon cDNA Library – Round 9 – Plaque A3

TTGGGGGTCTGGGAC**GAATTCA**AGCCAGCAGTACAACTATGCCTACCCCTACAGCTACTACTATCCCATGAGC
ATGTACCAGAGCTATGGCTCCCCTTCCCAGTATGGGATGGCCGGCTCCTATGGCTCAGCCACACCCCAGCAGC
CATCCGCACCTGACCTCAAGGAGAGTGGGGCAGCTGTGTGAGTCCCACATCCTGGGCAGAGGGCCTGGTGGGG
CCCCTTGCTAGGAGAAGGGAAGACGCCCGAGACGCTGCTTCCCCAGAAGTGCTGGGGCAGGGAGGCCCAGGAG
ATGAGAGAGAAGGTCCGAGTAGGTGATAGAAGACAAGGGGGAGACCGAGCCGGAGGCTGAGGAAAGGAAGAGG
GCACGGAGTTGCCAGGAGCAAACCAAAGTGAAGAGAGAGATAGGAAGCTGCCTCGGGGCCACCCCTCAAAGGG
GGTGTGTCCCACAAACGCTGCTATGGGTGGGGTGGGGGGCTGGGGTGCTGCGTAGCCAGTGTTTGACTTTCTT
TTCAAGTGGGGGAAAGTGGGAGAGGACTGAGAGTGAGGCAAGTTCTCCCCAGCCCCTGTCCGTCTGTCTGTCT
GTCTGTGGTGGTTTCTGTTTCTTGGGAGGCATGGTAGGATCATAAGTCATTCCCCTCCCCTTCCAGGCCTCCT
GCTATATTTGGGGGACCTGACTGGTTTGGCTGGAGTCCCATGAGGATGTGGGCCCTTTAATAAAGGATAGCAA
ACAGGGAAAAAAAAAAAAAAAAAAAAAAAGCTTGGCGGCCCCCCCCCAAGTAATTATTAACCCCCTGGGGCCCCC
TAACCGGTCTTTGGAGGTGTTTAA (up primer)

LOCUS NM_052925 3970 bp mRNA linear PRI 15-AUG-2011
DEFINITION **Homo sapiens leukocyte receptor cluster (LRC) member 8 (LENG8), mRNA.**
ACCESSION NM_052925 VERSION NM_052925.2 GI:215820642

```
   1 ccactgtact ctatcgtggg cgacgacaga gcaagagcaa gactccgtct ccgagaacaa
  61 caacaacagc aacaagaaaa caacaataaa aaaaataagg ctgcgtggga ggcagaaaga
 121 gctaatgcgg ccacgcttgt cccctcgggg ccaccgtccc cacccagact tccggtctgc
 181 cttaaaatgt tcatgcgtaa gtgcgtgggc aggaaggcgg gctcaagcgc agctcgtggc
 241 gttcattggc tgtgcagggc cgagggaggc ggtgcaaggc cgccgcgtga cgtcaggacg
 301 ccgcggtcag gacgtcgaag ccaaagaaga ccagagccag ccgggtggca cagcggtgtc
 361 gtggccgtgt tgctgatcgc ctgggtggtt gttggcgtgt ccctgcagcg aaggatcctg
 421 gttggcagtg aaaaagcagt ctggctcccg aggtccaccc cttatacccc aaggtccaga
 481 tggcggccaa cgtgggtgat caacgtagca cagattggtc ttctcagtac agcatggtgg
  //
 661 ccagcagcaa tgggcctgtg gccagtgcac agtacgtgtc ccaggcagaa gcctcagctt
 721 tgcagcagca gcagtactac cagtggtacc agcagtacaa ctatgcctac ccctacagct
 781 actactatcc catgagcatg taccagagct atggctcccc ttcccagtat gggatggccg
 841 gctcctatgg ctcagccaca ccccagcagc catccgcacc ccaacaccaa gggactctga
 901 accagccccc agtccccggc atggatgaga gcatgtccta ccaggctccc cctcagcagc
  //
2821 cgggcccgga caactccagc atcgactgcc gcctcagcct ggcgcagctg tcagccttct
2881 gagcacccag cgaggagggg cgggggcagg ggctgcagcc cccagcgctg cctttgcgga
2941 ttctgttttt gagccgtgga cttgggttgt aaatttattt gtggggagtg cgctccagga
  //
3901 cctttccttg gagcactgag caccatttgg aagcttgaga gaaaccaaaa ttaaagagag
3961 aaagagagag //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame):
Frame 1: **SQQYNYAYPYSYYYPMSMYQSYGSPSQYGMAGSYGSATPQQPSAPDLKESGAAV**
Frame 2: ASSTTMPTPTATTIP
Frame 3: PAVQLCLPLQLLLSHEHVPELWLPFPVWDGRLLWLSHTPAAIRT

BLASTp and alignment with full protein from Leukocyte receptor cluster:

```
maanvgdqrs tdwssqysmv agagrengme tpmhenpewe karqalasis ksgaaggsak
sssngpvasa qyvsqaeasa lqqqqyyqwy qqynyaypys yyypmsmyqs ygspsqygma
gsygsatpqq psapqhqgtl nqppvpgmde smsyqappqq lpsaqppqps npphgahtln
sgpqpgtapa tqhsqagpat gqaygphtyt epakpkkgqq lwnrmkpapg tgglkfniqk
rpfavttqsf gsnaegqhsg fgpqpnpekv qnhsgssarg nlsgkpddwp qdmkeyverc
ftaceseedk drtekllkev lqarlqdgsa ytidwsrepl pgltrepvae spkkkrweaa
```

sslhpprgag satrgggaps qrgtpgagga grargnsftk fgnrnvfmkd nssssstdsr
srsssrsptr hfrrsdshsd sdssysgnec hpvgrrnppp kgrggrgahm drgrgraqrg
krhdlaptkr srkkmaalec edperelkkq kraarfqhgh srrlrleplv lqmsslessg
adpdwqelqi vgtcpditkh ylrltcapdp stvrpvavlk kslcmvkchw kekqdyafac
eqmksirqdl tvqgirteft vevyethari alekgdheef nqcqtqlksl yaenlpgnvg
eftayrilyy iftknsgdit telayltrel kadpcvahal alrtawalgn yhrffrlych
apcmsgylvd kfadrerkva lkamiktfrp alpvsylqae lafegeaacr afleplglay
tgpdnssidc rlslaqlsaf

First part of cDNA insert matches with:

LOCUS NM_006612 7935 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens kinesin family member 1C (KIF1C), mRNA.**
ACCESSION NM_006612 XM_375394 VERSION NM_006612.5 GI:291327508

1 ccagctcgcg ctgcccgggc gggcgccggc cgctggcgcc gctactgctg ccgcccccgg
//
301 agccccccctg gtctggggcc aggacgccag ctgaggaggg caggagtgtc tggagct**atg**
361 **gctggtgcct cggtgaaagt ggcagtgagg gttcggccct ttaacgcccg tgagaccagc**
//
3541 **ccccagccac cccaacccta cccagcccag cggcccccag ggccccgcta ccccccatac**
3601 **actactcccc cacgaatgag acggcagcgt tctgcccctg acctcaagga gagtggggca**
3661 **gctgtgtga**g tcccacatcc tgggcagagg gcctggtggg gccccttgct aggagaaggg
3721 aagacgcccg agacgctgct tccccagaag tgctggggca gggaggccca ggagatgaga
//
4141 tgctatattt gggggacctg actggtttgg ctggagtccc atgaggatgt gggcccttta
4201 ataaaggata gcaaacaggg agcttgtggc ctgtttgttt tgggttttca tggaggtgta
4261 ggttatataa ggcaatggca caggtcttaa gcatacttat cagtgaagta ttgtatgtgt
//
7861 tttgttgttg ttgtcacctg gaacttttgt atcttgaata aatttgggga tcaaataaaa
7921 aaaaaaaaaa aaaaa //

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame):

Frame 1: **SQQYNYAYPYSYYYPMSMYQSYGSPSQYGMAGSYGSATPQQPSAPDLKESGAAV**
Frame 2: ASSTTMPTPTATTIP
Frame 3: PAVQLCLPLQLLLSHEHVPELWLPFPVWDGRLLWLSHTPAAIRT

Colon cDNA Library – Round 9 – Plaque D1

ATGTCCGGGGCCGGGAC**GAATTCT**GAGTCTGTTGACCAAGGTGCCAAACCAGAGAGTAAATCAGAACCTGTAG
TTTCCACTCGGAAAAGACCAGAGACCAAACCTTCCAGTGACCTTGAGACTTCAAAAGTTCTCCCTATTCAGGA
TAATGTTTCCAAAGATGTACCCCAGACCAGATGGGGTTATTGGGGGAGCTGGGGCAAGTCCATACTCTCCTCA
GCCTCGGCTACAGTAGCTACAGTAGGACAAGGCATTTCAAATGTCATCGAGAAGGCAGAGACTTCCCTTGGAA
TCACCAAAAGCCTGAGGAGGCCAGAGACCTCTTACCGGCAAACTGGTCTAGCTCCTGGGCAAGAGTATGAGAT
ATCTCTGCACATAGTGAAAAACAATACCCGGGGCCCTGGCCTGAAGAGGGTGACCACCACACGCTTGGATGCC
CCCAGCCAGATCGAGGTGAAAGATGTCACAGACACCACTGCCTTGATCACCTGGTTCAAGCCCCTGGCTGAGA
TCGATGGCATTGAGCTGACCTACGGCATCAAAGACGTGCCAGGAGACCGTACCACCATCGATCTCACAGAGGA
CGAGAACCAGTACTCCATCGGGAACCTGAAGCCTGACACTGAGTACGAGGTGTCCCTCATCTCCCGCAGAGGT
GACATGTCAAGCAACCCAGCCAAAGAGACCTTCACAACAGGCCTCGATGCTCCCAGGAATCTTCGACGTGTTT
CCCAGACAGATAACAGCATCACCCTGGAATGGAGGAATGGCAAGGCAGCTATTGACAGTTACAGAATTAAGCT
TGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGTCTTGGGAGGGGGTTACTA (up
primer)

LOCUS NM_018691 2886 bp mRNA linear PRI 10-MAR-2011
DEFINITION **Homo sapiens family with sequence similarity 114, member A2 (FAM114A2), mRNA.**

ACCESSION NM_018691 VERSION NM_018691.2 GI:93102374

CDS 96..1613
1 tctgcgctga aggagcactt ccggggcagt aggaacgccg agtcggctgc cgtggctgtg
61 ctgagggtgg cggccggata gctgatgttc taatc**atgtc agataaagat gatattgaga**
121 **ctccactgct aactgaagca gcccccatcc ttgaagatgg aaactgtgag ccagccaaga**
181 **attctgagtc tgttgaccaa ggtgccaaac cagagagtaa atcagaacct gtagtttcca**
241 **ctcggaaaag accagagacc aaaccttcca gtgaccttga gacttcaaaa gttctcccta**
301 **ttcaggataa tgtttccaaa gatgtacccc agaccagatg gggttattgg gggagctggg**
361 **gcaagtccat actctcctca gcctcggcta cagtagctac agtaggacaa ggcatttcaa**
421 **atgtcatcga gaaggcagag acttcccttg gaatccctgg tcccagtgaa atttcaactg**
481 **aagtcaagta tgtagcagga gagacaaatg ccaaagagaa tgaaaactcc tccccagtgg**
//
1501 **tccaggacgc ctttcagcta ctcttacctg tgctagagat ctctctcatt gagaacaaga**
1561 **ttgaatcaca cagacatgag ctgcagggcc agaaaccttt gttagaacat tga**agaatgg
1621 agacgttttg acctgggact tgtgacggcc aaggaatgcc accttattct ggctactcct
//
2821 gttatttgat cctttcattt acttttgtaa acacgagccc taatatctta cttagaccaa
2881 aaaaaa //

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **ESVDQGAKPESKSEPVVSTRKRPETKPSSDLETSKVLPIQDNVSKDVPQTRWGYWGS**
WGKSILSSASATVATVGQGISNVIEKAETSLGITKSLRRPETSYRQTGLAPGQEYEI
SLHIVKNNTRGPGLKRVTTTRLDAPSQIEVKDVTDTTALITWFKPLAEIDGIELTYG
IKDVPGDRTTIDLTEDENQYSIGNLKPDTEYEVSLISRRGDMSSNPAKETFTTGLDA
PRNLRRVSQTDNSITLEWRNGKAAIDSYRIKLAAALE
Frame 2: SLLTKVPNQRVNQNL
Frame 3: VC

Also matched tenascin receptor:

LOCUS NM_002160 8605 bp mRNA linear PRI 21-AUG-2011
DEFINITION **Homo sapiens tenascin C (TNC), mRNA.**
ACCESSION NM_002160 VERSION NM_002160.3 GI:340745336

CDS 413-7018
1 aattcgccaa ctgaaaaagt gggaaaggat gtctggaggc gaggcgtccc attacagagg
//
301 gggacaactg atttgaagtc tactctgtgc ttctaaatcc ccaattctgc tgaaagtgag
361 ataccctaga gccctagagc cccagcagca cccagccaaa cccacctcca cc**atgggggc**
421 **catgactcag ctgttggcag gtgtctttct tgctttcctt gccctcgcta ccgaaggtgg**
//
2581 **caaggagaca tctgtggaag tggagtggga tcctctagac attgcttttg aaacctggga**
2641 **gatcatcttc cggaatatga ataaagaaga tgagggagag atcaccaaaa gcctgaggag**
2701 **gccagagacc tcttaccggc aaactggtct agctcctggg caagagtatg agatatctct**
//
3121 **ccagacagat aacagcatca ccctggaatg gaggaatggc aaggcagcta ttgacagtta**
3181 **cagaattaag tatgccccca tctctggagg ggaccacgct gaggttgatg ttccaaagag**
3241 **ccaacaagcc acaaccaaaa ccacactcac aggtctgagg ccgggaactg aatatgggat**
//
6901 **cagtcagggc gttaactggt tccactggaa gggccacgaa cactcaatcc agtttgctga**
6961 **gatgaagctg agaccaagca acttcagaaa tcttgaaggc aggcgcaaac gggcataa**at
7021 tccagggacc actgggtgag agaggaataa ggcccagagc gaggaaagga ttttaccaaa
//
8521 atgagatgga ctggggttca tagaaaccca atgacttgat tgtggctact actcaataaa
8581 taatagaatt tggatttaaa aaaaa //

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **ESVDQGAKPESKSEPVVSTRKRPETKPSSDLETSKVLPIQDNVSKDVPQTRWGYWGS**
WGKSILSSASATVATVGQGISNVIEKAETSLGITKSLRRPETSYRQTGLAPGQEYEI
SLHIVKNNTRGPGLKRVTTTRLDAPSQIEVKDVTDTTALITWFKPLAEIDGIELTYG
IKDVPGDRTTIDLTEDENQYSIGNLKPDTEYEVSLISRRGDMSSNPAKETFTTGLDA
PRNLRRVSQTDNSITLEWRNGKAAIDSYRIKLAAALE
Frame 2: SLLTKVPNQRVNQNL
Frame 3: VC

BLASTp on frame 1 protein:
Tenascin precursor protein matches query 90-261 aa

8.3.3.3.2 Human Alzheimer's brain

Alzheimer's brain cDNA Library – Round 15 – Plaque B4

CAAAGCCGCCCGATCCGCGAC**GAATTCT**GAGGTTGGGAGTGAACAGAGAGTAATGTGGATTCTGAATTTGCAC
GGACTCAGTTTGAGGATCTTGTTCCCTCACCAACCTCTGAAAAGGCTTTTCTAGCGCAAATCCATGCCCGAAA
ACCTGGGTACATTCACAGTGGAGCTACCACAAGTACCATGCGTGGCGACATGGTTACGGAGGATGCAGATCCC
TATGTGCAGCCTGAAGATGAAAACTATGAAAATGACTCTGTCCGGCAGCTGGAGAATGAGCTCCAGATGGAGG
AATACCTGAAACAGAAGCTGCAAGATGAAGCTTATCAGGTCAGCTTGCAAGGTTGAATGCAATATCCTTTTAT
CACACTCCTCTCAACTCCACGTCAGCACTGAGACCAGACTCGAGCACCCCTGTCCTGTAAGCGAGACAAAATG
GCGTGTGTTATTTTGGGGTTTTGTGTTTTTTGGTGGGTTTCTTTCCTTGGCTCTCCAGATTTACTTTTGGGGC
CTGTTCTAAGTGCAAACCCAGCAAGTTTCACTTGTCCTGTCCATTAGATACAACTACATCTTGCGGGGGTTGT
TTCTTTCTTGTTCCACAATGAATTGCACATCCATCTCCATCAGCTTGATAGCCTGTTAATAAGCACTGGTCTA
ACACAGCCAACCCTCCTCCACAGCGCCATATTAATGGAGGAGGGGAGGAAGGTGAAATCTACTGCATGGGATT
CAGGAAACAGTTGTGGTTGGTCAGGACGGAAGTTGGGGTAAGTTTGGTTGGTCAGAGGGAGTTGTGCTGGAGA
TTGTGAAAAATGGGTTCTTGAATGATCTACTATAAGCAGGAAAGGTTCATTTGTAAGTAGTAATGTGAACTGA
ATTGCATTAAGAGTGTGTGGGCCTTTGTTGTGATATACTATGTATTTTCTTATATGCATGAAGCCAAACTGTT
GCATCA (up primer)

GCGGAAGGTCTACTGTTCTCGAGTGCGGCCGCAGCTTTTTTAGTGTTTTTAAAAACATTTTTTAATGAAGAAA
TTATTCCACTCCAAAATATAGACACACAATTAAATAGTTTAATCACTTCAAAATATAACATGTCCCCAGGAGG
TTCCAACACACACACACACACACACACACAAAACAATACTTAACAGCAATACAAAAATCCATACTATAGTATT
TTTTTTTTACAATACCAATGAATATTGTGAAATCTTTAGTGTCTCTGAATTGATTATTTCAAAGGGTGAAAAA
TTGCAATAGTCATAATTAAAAATGTCTTACAGTTAATTCCCCTCAAGATGCAACACTTCTGATTTTTTTTTCT
CTTTTTGGTTCAAAAGTAAAAATGCCCTTTCTTCTAAGGACAGTGGAGTTCTCACTTGGGGCTCCTTGCGGGG
AATCAAACCGGGGGGAATGCAGCAACCAGGGGTTGCTGGATTTACAAGATCTATCTACGGTTAACACCCAAAA
ACTAATAAGGGT (down primer)

LOCUS NM_001198944 5741 bp mRNA linear PRI 29-APR-2012
DEFINITION **Homo sapiens dystrobrevin, alpha (DTNA), transcript variant 16, mRNA.**
ACCESSION NM_001198944 VERSION NM_001198944.1 GI:312147281

CDS 444-1631
```
   1 ggtggttctc atagcaacca gcaccaaatg aaagagtaca cgtcatgggt aaggcaaggt
  61 ccaggcaata cttggagttg gatgtgtgag tgttgcccag agttaaaggg aatatcaagg
 121 atggtggcag aattaaaggg caaacttacc ttccatccca ggtctagtat tcagtccccc
 181 ttcctcccac tctccaccct actgcgtgct cttacttatt ctgttggaac ccagtgtagc
 241 ttcctgagat ttaaaatata ggaagactca aaactatgtt acttcttgga atatagataa
 301 ttcagtactg atttaaataa caactgaata gcaaggacct tctggaacat aatcttacat
 361 tcacaaaatc acctgctaag aagctgacta atgcattaag caagtccctg agctgtgctt
 421 ccagccgtga acctttgcac ccc
```
atgttcc cagatcagcc tgagaagcca ctcaacttgg
ctcacatcgt gcctcccaga cctgtaacca gcatgaacga caccctgttc tcccactctg

```
  //
1201 tacaagaggc atttgcacaa agttcaagaa gaaacttaag gaatgacttg ctagtggctg
1261 cagattccat cactaacact atgtcctctc ttgtgaaaga gctgaattct gaggttggga
1321 gtgaaacaga gagtaatgtg gattctgaat ttgcacggac tcagtttgag gatcttgttc
1381 cctcaccaac ctctgaaaag gcttttctag cgcaaatcca tgcccgaaaa cctgggtaca
1441 ttcacagtgg agctaccaca agtaccatgc gtggcgacat ggttacggag gatgcagatc
1501 cctatgtgca gcctgaagat gaaaactatg aaaatgactc tgtccggcag ctggagaatg
1561 agctccagat ggaggaatac ctgaaacaga agctgcaaga tgaagcttat caggtcagct
1621 tgcaaggttg aatgcaatat ccttttatca cactcctctc aagcaagcta tagtcagaca
1681 gatgatgaaa gtaacatgaa aactggtgat gatggagagc cctgtggcca cacagaggag
1741 gaagacagca gcctggcagc agcctcaccg aagagaggaa ccaccacacc cagcagctcc
  //
5641 cactcaattc tttgaatcct ctgcatctag ccatgtattc tgcaaatatt aagtgctcaa
5701 tggttttttt gttgaattac tgaataaatg aattagtggt g //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SEVGSETESNVDSEFARTQFEDLVPSPTSEKAFLAQIHARKPGYIHSGATTSTMRGDM VTEDADPYVQPEDENYENDSVRQLENELQMEEYLKQKLQDEAYQVSLQG**
Frame 2: LRLGVKQRVMWILNLHGLSLRILFPHQPLKRLF
Frame 3: -

Alzheimer's brain cDNA Library – Round 15 – Plaque B5

GCGTCTGTCGAACTCGAC**GAATTCA**AGCGATCAACGTAGCACAGATTGGTCTTCTCAGTACAGCATGGTGGCT
GGGGCAGGCCGAGAGAATGGCATGGAGACGCCGATGCACGAGAACCCGGAGTGGGAGAAGGCCCGTCAGGCCC
TGGCCAGCATCAGCAAGTCAGGAGCTGCCGGCGGCTCTGCCAAGTCCAGCAGCAATGGGCCTGTGGCCAGTGC
ACAGTACGTGTCCCAGGCAGAAGCCTCAGCTTTGCAGCAGCAGCAGTACTACCAGTGGTACCAGCAGTACAAC
TATGCCTACCCCTACAGCTACTACTATCCCATGAGCATGTACCAGAGCTATGGCTCCCCTTCCCAGTATGGGA
TGGCCGGCTCCTATGGCTCAGCCACACCCCAGCAGCCATCCGCACCCCAACACCAAGGGACTCTGAACCAGCC
CCCAGTCCCCGGCATGGATGAGAGCATGTCCTACCAGGCTCCCCCTCAGCAGCTGCCGTCGGCTCAGCCCAGT
CAGGAGGCCTGGAGGCTTGGGGAGAGGTGTGGAGAAGGAGAGAACATGGCCCAGGCCCTTTCCTTCCCCCTGT
GCTGACAGCATTGCTGTGGGGGTGGCCCACTGCCCTCCCCTGGCCCTCATGTCCCCCCGGGGCTGGGGTCCGC
CTGCCTGTGCTGTGCTTGCGACGTGCATCAATAAACCACCATGGCCTGAGGGCCCTGCTCGTCAAAAAAAAAA
AAAAAAAAAAAAAAACTTGGGGCCCCCCCTCAAAAAATTATTACACCCCTGGGGCCTCCAAACGGGTTTTGGG
GGGGGGGTTTA (up primer)

TCCAATATCTCCTCTGTTCTCGAGTGCGGCCGCAAGCTTTTTTTTTTTTTTTTTTTTTTTAAAAAAAGGGC
CCCCCTGGCAAGGGGGTTTTTTTAATCCCCTTCCAAAGAACACCCCAGGGGGGGGGCCCCCCCCCCCGGGGGA
ACATAAGGCCCGGGGAAGGGAAGGGGCCCCCCCCCCCCAAAAATTTTTTCACCAAGGGGGAAAAAAAAGGGCCG
GGGCCTTTTTTTCCCTTTCCCCAACCCTTCCCCAAACCCCCAGGGCCCCTTAAGGGGGTTAACCAAACGGAAT
TTTTTAAGGGGAACCCGGGAAGAAAATTTTCCCTCCCTTCCGGGGAAGGGGGGGTGGGTTAAAATCCCTTTGG
TTTTGGGGGGGAAAAGGCTCCGGGGGGGGGGTAAACCAAAAAAACCGGGCCTCCCCATATGGGAAGGGGAACC
CAAACTTTTGGAAATTGCCCTGGGAAAAAAAAAACTTGAGGGGGAGGGAAAATTTTAACGGGGGAACCCCGGG
GAAAAACGGCT (down primer; sequencing trace is contaminated, more than one sequence present)

LOCUS NM_052925 3970 bp mRNA linear PRI 26-MAR-2012
DEFINITION **Homo sapiens leukocyte receptor cluster (LRC) member 8 (LENG8), mRNA.**
ACCESSION NM_052925 VERSION NM_052925.2 GI:215820642

CDS 480-2882
```
   1 ccactgtact ctatcgtggg cgacgacaga gcaagagcaa gactccgtct ccgagaacaa
  61 caacaacagc aacaagaaaa caacaataaa aaaaataagg ctgcgtggga ggcagaaaga
 121 gctaatgcgg ccacgcttgt cccctcgggg ccaccgtccc cacccagact tccggtctgc
```

```
 181 cttaaaatgt tcatgcgtaa gtgcgtgggc aggaaggcgg gctcaagcgc agctcgtggc
 241 gttcattggc tgtgcagggc cgagggaggc ggtgcaaggc cgccgcgtga cgtcaggacg
 301 ccgcggtcag gacgtcgaag ccaaagaaga ccagagccag ccgggtggca cagcggtgtc
 361 gtggccgtgt tgctgatcgc ctgggtggtt gttggcgtgt ccctgcagcg aaggatcctg
 421 gttggcagtg aaaaagcagt ctggctcccg aggtccaccc cttataccca aaggtccaga
 481 tggcggccaa cgtgggtgat caacgtagca cagattggtc ttctcagtac agcatggtgg
 541 ctggggcagg ccgagagaat ggcatggaga cgccgatgca cgagaacccg gagtgggaga
 601 aggcccgtca ggccctggcc agcatcagca agtcaggagc tgccggcggc tctgccaagt
 661 ccagcagcaa tgggcctgtg gccagtgcac agtacgtgtc ccaggcagaa gcctcagctt
 721 tgcagcagca gcagtactac cagtggtacc agcagtacaa ctatgcctac ccctacagct
 781 actactatcc catgagcatg taccagagct atggctcccc ttcccagtat gggatggccg
 841 gctcctatgg ctcagccaca ccccagcagc catccgcacc ccaacaccaa gggactctga
 901 accagccccc agtccccggc atggatgaga gcatgtccta ccaggctccc cctcagcagc
 961 tgccgtcggc tcagcccct cagccctcaa atcccccaca tggggctcac acgctgaaca
1021 gtggccctca gcctgggaca gctccagcca cacagcacag ccaggcgggg cccgccacgg
1081 gccaggccta tgggccacac acctacaccg aacctgccaa gcccaagaag ggccaacagc
  //
2761 tggccttcga gggcgaggcc gcctgccggg ccttcctaga gcccctgggc ctggcctaca
2821 cgggcccgga caactccagc atcgactgcc gcctcagcct ggcgcagctg tcagccttct
2881 gagcacccag cgaggagggg cgggggcagg ggctgcagcc cccagcgctg cctttgcgga
2941 ttctgttttt gagccgtgga cttgggttgt aaatttattt gtggggagtg cgctccagga
3001 agagccacca tccctgcccc cgttttccca ccggggagtc tgtacagaga tttttctacg
  //
3841 ccagatgttc ctcctctgcc tccccttccc ctcctctccc ctccttttcc ttccttcctt
3901 cctttccttg gagcactgag caccatttgg aagcttgaga gaaaccaaaa ttaaagagag
3961 aaagagagag
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SSDQRSTDWSSQYSMVAGAGRENGMETPMHENPEWEKARQALASISKSGAAGGSAKS SSNGPVASAQYVSQAEASALQQQQYYQWYQQYNYAYPYSYYYPMSMYQSYGSPSQYG MAGSYGSATPQQPSAPQHQGTLNQPPVPGMDESMSYQAPPQQLPSAQPSQEAWRLGE RCGEGENMAQALSFPLC**
Frame 2: QAINVAQIGLLSTAWWLGQAERMAWRRRCTRTRSGRRPVRPWPASASQELPAALPSP AAMGLWPVHSTCPRQKPQLCSSSSTTSGTSSTTMPTPTATTIP
Frame 3: KRST

Alzheimer's brain cDNA Library – Round 15 – Plaque C6

```
TAAAGTGAaTTCGAATCCGCGAC GAATTCA GCGAGCTGTAGGGGTTAGTTTGATTAGTCATTGTTGGGGGTTC
TTGGCTGGGAGGTCCGAGGTGTCCCAGCTTCCCCCTAGGCAGAGCCCCTGCAGGTCTTATTTAGGGAAACCCC
CACACGCAGAGGGGCATTTGTCTCAGTGGGTTGAAAAGGCTACTTTCCTCCAGCTCATTGATCAGATCCCAGA
GAACAGTGGGGAGCCTGAGCCCAGGGGCCGAGGCATCTTGAGGGGCTGGGGCTTTCTGAGAGCCTTTATCTCA
GCACAGGAGATAAGACTCCAGTGGCTGCAGGAGCCATAAAGCAAGGAGGGGCTGGCCGTTGAGGATCCCGGGG
AGCCGGGCTTTGAACTCTGCCCTCATGGGGGTGGGAAGCCTTTATGGCTGGGCTGCTCTGGGGCCGAGGTGGG
AGCAGAGCCATCACGGGAGAAAAGAGTCCCCAGAGCCGGCCCCCTGGCACCCAGACTGCACGGTGAGACAACG
GAGCCCGGAGTGGTGGGACCAGCGGCGGAAACAGCGCTCCATCCGGGCTTTATTGACTTGTAAACGGTACCAT
TAAAGAGGAAGGCCTCACCTTGGCTGGAAGCAGCTGTGTGCGCTGCCGTAAATCTAGGGCGAGCTCTGAGGCT
GCAGCCGGGCCTGGCAAATGCCAAGCATGCCCTGGGCAGAAGGGGGTGCAGGCCAGGGATCCCCACCACACCC
AGCTCCTAGCTCCACTGAGGGGCTGAGTATGGGAATTTATAAGAAGGATGTTTACCAGGGGAAATATCAGTGG
CTAATTAGACTTTGCAAATAAAAAAACTTTTCAAAAAAAAAAAAAAAAGCTTGCGGCCGCACTCGAGTAACTAG
TTAACCCCTTGGGGCCTCTAACTCT (up primer)

TAACCAACGCTCTAGTTCTCGAGTGCGGCCGCAAGCTTTTTTTTTTTTTTTGAAAATTTTTTTTATTTGCAAA
GTCTATTTAGCCACTGATTTTTCCCCTGGAAAACATCCTTTTTATAATTTCCCATACTCACCCCCTCAGTGGA
GTTAGGAGCTGGGGGGGGGGGGGATCCCGGGCCTGCCCCCCCTTTTGCCCAGGGCATGTTTGGCATTTGCCAG
GCCCGGTTGCACCCTCAAAGCTCGCCCTAAATTTACGGCAGCGCACACAGTTGTTTCCAGCCAAGGGGAGGCC
```

TTCCTTTTTAATGGTACCGTTTACAAGTCAATAAAGCCCGGAGGGAGCGTTTTTTCCCCCCCTGGTCCCACCA
CTCCGGGTTCCGTTGTCTCACCGGGCAGTCTGGGTGCCAGGGGGCCGGTTCTGGGGACTTTTTTCTCCCGTGA
TGGCTCTGCTCCCACCTCGGCCCCAAAGCAGCCCACCCATAAAGGTTTCCCACCCCCATGAGGGCAAATTTCA
AAGCCCGGCTCCCCGGGATCCTCAACGGCCACCCCCTCCTTGTTTTATGGTTCCTGCAGCCACTGGAGTCTTA
TCTCCTGTGCTGAAATAAAGGCTCTCAAAAAGCCCCAGCCCCTCAAAATGCCTCGGCCCCTGGGCTCAGGCTC
CCCATTGTTCTCTGGAATCTGATCAATGACCTGGAGGAAAGTACCCTTTTCAACCCATTGAAACAAATGCCCC
CTCTGCGTGTGGGGGTTTCCCTAAATAAAACCTGCAGGGCTCTGCCTAGGGGAAGCTGGAACACCTCGAACCT
CCCAGCCAAGAACCCCCAACAATGACTAATCAAACTAACCCCTACAGCTTCGCTTGATTTCGATCCCCGAGTA
TCACACCTGACTGGATACCGAAAGCTTCAT (down primer)

LOCUS NW_001838766 3466 bp DNA linear CON 29-JUL-2011
DEFINITION **Homo sapiens chromosome 2 genomic contig, alternate assembly HuRef SCAF 1103279187422, whole genome shotgun sequence.**
ACCESSION **NW_001838766** REGION: 8179151..8182616
VERSION NW_001838766.1 GI:157696327

Query pos. 35 - 809 match DNA sequence pos. 11124120 - 11124893

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: SNPRRFSEL
Frame 2: RIRDDSASCRG
Frame 3: ESATIQRAVGVSLISHCWGFLAGRSEVSQLPPRQSPCRSYLGKPPHAEGHLSQW
VEKATFLQLIDQIPENSGEPEPRGRGILRGWGFLRAFISAQEIRLQWLQEP

BLASTp of frame 3:
no significant similarity (<40 aa match 105aa sequence)

Alzheimer's brain cDNA Library – Round 15 – Plaque F4

CCGTTACCCGAAAATGGGAC**GAATTCC**AACAAAAGCCTCTCTGGTGAACTCAGAGAAGTTAAAAAGGACCTTA
GTCAGTTACAAGAAAACTTAAAGATTTCGGAAAACATGAATTTACAATTTGAAAACCAAATGAATAAGACAAT
CAGAAACTTATCTACGGTAATGGATGAAATCCACACTGTTCTCAAGAAGGATAATGTAAAGAATGAAGACAAA
GATCAAAAATCCAAGGAGAATGGTGCCAGTGTATGATAAAATCCATGTAGTGATGAGGAATGGTGTTAAATAA
TGTAATATATAAAAATCATGATATAAGAATGTTTGAAGGTGATGCATGTTTGATTTTAGTAGTATAAATGTAT
TTTAGTTCAAATGATGTATAAAGTTTTATGAATGTGAGTTTCTGCTTTTGAAAATTGCTTGTAATTCCTAGCC
TTCAAATTATTAAACACTCCTTGAGTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGTTTGGGGCCCC
CCCCCAAAAAAATTTTAACCCCTTGGGGCCCTAAAAGGGGTTTTGGGGGGTTTAAAA (up primer)

LOCUS NM_018237 3858 bp mRNA linear PRI 15-APR-2012
DEFINITION **Homo sapiens cell division cycle and apoptosis regulator 1 (CCAR1), mRNA.**
ACCESSION NM_018237 VERSION NM_018237.2 GI:46852387

CDS 120-3572
```
   1 gaagttggcg catgcgccta aagctgacgg gtttgaaatg gcttcgatgt tagccgggac
  61 ccgactcaga tcgatgctat agaagacaaa caagggaagg ttttttttcc ttttgcatca
 121 tggctcaatt tggaggacag aagaatccgc catgggctac tcagtttaca gccactgcag
 181 tatcacagcc agctgcactg ggtgttcaac agccatcact ccttggagca tctcctacca
  //
3181 atgttggcct cattgtgtac aatggtgcaa tggtagatgt aggaagcctc ttgcaaaaat
3241 tggaaaagag cgaaaaagta agagctgagg tagaacagaa gctgcagtta ctagaagaaa
3301 aaacagatga agatgaaaaa accatattaa atttggagaa ttccaacaaa agcctctctg
3361 gtgaactcag agaagttaaa aaggacctta gtcagttaca agaaaactta aagatttcgg
3421 aaaacatgaa tttacaattt gaaaaccaaa tgaataagac aatcagaaac ttatctacgg
3481 taatggatga aatccacact gttctcaaga aggataatgt aaagaatgaa gacaaagatc
3541 aaaaatccaa ggagaatggt gccagtgtat gataaaatcc atgtagtgat gaggaatggt
3601 gttaaataat gtaatatata aaaatcatga tataagaatg tttgaaggtg atgcatgttt
3661 gattttagta gtataaatgt attttagttc aaatgatgta taaagtttta tgaatgtgag
```

3721 tttctgcttt tgaaaattgc ttgtaattcc tagccttcaa attattaaac actccttgag
3781 tgaaataatt ttgcattgca aagtgtttta ggatgaactt tgttatagtt ttaactccaa
3841 taaagttcat cagtttaa //

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SNKSLSGELREVKKDLSQLQENLKISENMNLQFENQMNKTIRNLSTVMDEIHTVLKK DNVKNEDKDQKSKENGASV**
Frame 2: PTKASLVNSEKLKRTLVSYKKT
Frame 3: QQKPLW

Alzheimer's brain cDNA Library – Round 15 – Plaque H5

AACCCGGCCGGAGGCCGGGAC**GAATTCA**GCAAATCCAGGTCGCGATCGAAGAGTCCCCCCAAGTCTCCTGAAG
AGGAAGGAGCGGTGTCCTCTTAAGAAAATGCTGCGGTCTCCTGTTTGATAAAAGAATATTGGCCAGTATTGCA
GATTTTAACTGATTTGGCTGATCCTCCAGGGACCAGTTTCTGTGGGCGTGTATTGGAGCAGATGTATCGGCAA
GCAGTGTAAACGGAGGACTTGGGGAAAAAGGACCACATAGTCCATCGAAGAAGAGTCCTTGGAACAAGCAACT
GGCTATTGAAAAGGTTATTTTGTAACATTTGTCTAACTTTTTACTTGTTTAAGCTTGCGGCCGCACTCGAGTA
ACTAGTTAACCCCTTGGGGCCTCTAAACGGGTTTGGAGGGGTTAAACTAGTTACTCGAGTGGGCGCAAGCTTA
AACAAGTAAAAAGTTAGACAAATGTTACAAAATAACCTTTTCAATAGCCAGTTGCTTGTTCAAGGACTCTTCT
TCGATGGACTATGTGGTCTTTTTCCCAAGTCCTCCGTTTACACTGCTGCCGATACATCTGCTCAATACACGCC
CACGAAACCGGGTCCCGGAGGATCACCCAATTAGTTAAAACTGCATACTGGCCATTTTTTTTTATAAAAGGGA
AACCCGCATTTTTTAAAAGGACCCGCTCCTCCTCTTCGGAAAATTGGGGGGGAATTTTCATGCCACTGGAATT
GTTGGAATTCGAACCCCAAGATTCAACCTGAAGGGAAAAAAAAGCTCCAAA (up primer)

LOCUS NR_036608 2112 bp RNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens serine/arginine-rich splicing factor 2 (SRSF2), transcript variant 3, non-coding RNA.**
ACCESSION NR_036608 VERSION NR_036608.1 GI:306482647

CDS n.a.
1 agaaggtttc atttccgggt ggcgcgggcg ccattttgtg aggagcgata taaacgggcg
61 cagaggccgg ctgcccgccc agttgttact caggtgcgct agcctgcgga gcccgtccgt
//
781 ccagatctcg ttcgcggtcc aggtcccggt ctcggtccag gagtcctccc ccagtgtcca
841 agagggaatc caaatccagg tcgcgatcga agagtccccc caagtctcct gaagaggaag
901 gagcggtgtc ctcttaagaa aatgctgcgg tctcctgttt gataaaagaa tattggccag
961 tattgcagat tttaactgat ttggctgatc ctccagggac cagtttctgt gggcgtgtat
1021 tggagcagat gtatcggcaa gcagtgtaaa cggaggactt ggggaaaaag gaccacatag
1081 tccatcgaag aagagtcctt ggaacaagca actggctatt gaaaaggtta ttttgtaaca
1141 tttgtctaac tttttacttg tttaagcttt gcctcagttg gcaaacttca ttttatgtgc
1201 cattttgttg ctgttattca aatttcttgt aatttagtga ggtgaacgac ttcagatttc
1261 attattggat ttggatattt gaggtaaaat ttcattttgt tatatagtgc tgacttttt
1321 tgtttgaaat taaacagatt ggtaacctaa tttgtggcct cctgactttt aaggaaaacg
//
1981 ggtgtatctt tttgcttatg aatattctgt attataacca ttgtttctgt agtttaatta
2041 aaacattttc ttggtgttag cttttctcag aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
2101 aaaaaaaaaa aa

Bold = Gene coding sequence (n.a.)
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: SA**NPGRDRRVPPSLLKRKERCPLKKMLRSPV**

Frame 2: QQIQVAIEESPQVS
Frame 3: SKSRSRSKSPPKSPEEEGAVSS

BLASTp of protein expressed in frame 1:

LOCUS AAH66958 179 aa linear PRI 28-JUL-2005
DEFINITION SFRS2 protein [Homo sapiens].
ACCESSION AAH66958 VERSION AAH66958.1 GI:44890463

8.3.3.3.3 Human colon tumour

Colon Tumour cDNA Library – Round 9 – Plaque A4

CCCGTGGGGCAC**GAATTCA**AGCAAAAAGTCCGGGAAGCTGAAAGTCCCCGAATGGGTGGATACCGTCAAGCTG
GCCAAGCACAAAGAGCTTGCTCCCTACGATGAGAACTGGTTCTACACGCGAGCTGCTTCCACAGCGCGGCACC
TGTACCTCCGGGGTGGCGCTGGGGTTGGCTCCATGACCAAGATCTATGGGGGACGTCAGAGAAACGGCGTCAT
GCCCAGCCACTTCAGCCGAGGCTCCAAGAGTGTGGCCCGCCGGGTCCTCCAAGCCCTGGAGGGGCTGAAAATG
GTGGAAAAGGACCAAGATGGCGGCCGCAAACTGACACCTCAGGGACAAAGAGATCTGGACAGAATCGCCGGAC
AGGTGGCAGCTGCCAACAAGAAGCATTAGAACAAACCATGCTGGGTTAATAAATTGCCTCATTCGTAAAAAAA
AAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTTTGGAGGGGTTAA (up
primer)

LOCUS NM_001022 872 bp mRNA linear PRI 14-
AUG-2011
DEFINITION **Homo sapiens ribosomal protein S19 (RPS19), mRNA.** ACCESSION
NM_001022 VERSION NM_001022.3 GI:48255921

```
  1 gtactttcgc catcatagta ttctccacca ctgttccttc cagccacgaa cgacgcaaac
 //
301 ctacgcccga cttgtgcgcc cgggaaaccc cgtcgttccc tttcccctgg ctggcagcgc
361 ggaggccgca cgatgcctgg agttactgta aaagacgtga accagcagga gttcgtcaga
421 gctctggcag ccttcctcaa aaagtccggg aagctgaaag tccccgaatg ggtggatacc
481 gtcaagctgg ccaagcacaa agagcttgct ccctacgatg agaactggtt ctacacgcga
 //
721 caagatggcg gccgcaaact gacacctcag ggacaaagag atctggacag aatcgccgga
781 caggtggcag ctgccaacaa gaagcattag aacaaaccat gctgggttaa taaattgcct
841 cattcgtaaa aaaaaaaaaa aaaaaaaaaa aa
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame):

Frame 1: **SKKSGKLKVPEWVDTVKLAKHKELAPYDENWFYTRAASTARHLYLRGGAGVGSMTK**
IYGGRQRNGVMPSHFSRGSKSVARRVLQALEGLKMVEKDQDGGRKLTPQGQRDLDR
IAGQVAAANKKH
Frame 2: AKSPGS
Frame 3: QKVREAESPRMGGYRQAGQAQRACSLR

Colon Tumour cDNA Library – Round 9 – Plaque B4

GGGGGGGTCGGGGAC**GAATTCA**AGCCAGAATACGAGGATGGGGCCCAGGAGAACACCTTCTCCATGGACCCCC
AACTGGAGCGGCAGGTGGAGACCATTCGCAACCTGGTGGACTCATACGTGGCCATCATCAACAAGTCCATCCG
CGACCTCATGCCAAAGACCATCATGCACCTCATGATCAACAATACGAAGGCCTCCAGCAGCTCGTGGTGGATG
AAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTT (up
primer)

LOCUS NM_001005360 3684 bp mRNA linear PRI 25-SEP-2011
DEFINITION **Homo sapiens dynamin 2 (DNM2), transcript variant 1, mRNA.**
ACCESSION NM_001005360 VERSION NM_001005360.2 GI:299758519

```
   1 gaggtcgctc gggtcgggtg tcgcctgaga accggatgag gcggcgaccg tgaggccgag
 //
 121 ggggccaggt cgttgagggt cggcggcggg cgaggagcgc agggcgctcg ggccgggggc
 181 cgccggcgcc atgggcaacc gcgggatgga agagctgatc ccgctggtca acaaactgca
 241 ggacgccttc agctccatcg gccagagctg ccacctggac ctgccgcaga tcgctgtagt
  //
1981 caaggacctg cggcagatcg agctggcctg tgactcccag gaagacgtgg acagctggaa
2041 ggcctcgttc ctccgagctg gcgtctaccc cgagaaggac caggcagaaa acgaggatgg
2101 ggcccaggag aacaccttct ccatggaccc ccaactggag cggcaggtgg agaccattcg
2161 caacctggtg gactcatacg tggccatcat caacaagtcc atccgcgacc tcatgccaaa
2221 gaccatcatg cacctcatga tcaacaatac gaaggcttc atccaccacg agctgctggc
2281 ctacctatac tcctcggcag accagagcag cctcatggag gagtcggctg accaggcaca
  //
2761 gcccaccatt atccgcccag ccgagccatc cctgctcgac taggcctcga gggggggcgtg
2821 ctctcggggg ggcctcacgc acccgcggcg caggagcttc agtggtctgg ggccctccgc
  //
3601 aagtgcctgc actctgtatt ctattaataa actaaaataa agggaagacg ctgctggtgg
3661 ctgctgaaaa aaaaaaaaaa aaaa  //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame):
Frame 1: **EFKPENEDGAQENTFSMDPQLERQVETIRNLVDSYVAIINKSIRDLMPKTIMHLMI NNTKASSSSWWMKLAAALE**
Frame 2: NSSQKTRMGPRRTPSPWTPNWSGRWRPFATWWTHTWPSSTSPSATSCQRPSCTS
Frame 3: IQARKRGWGPGEHLLHGPPTGAAGGDHSQPGGLIRGHHQQVHPRPHAKDHHAPH DQQYEGLQQLVVDEACGRTRVTS

Colon Tumour cDNA Library – Round 9 – Plaque B5

GGATTTCTGAGCAC**GAATTCC**ATGTTTCAGGAGCCTAAGACCCTCCCAGAGCCCAGGGGCTTCACCGCAGACC
CCAAGCCATTGAGCACATCACCCAAAGCAGTGGCCAACATCGCGGACCCCTGTGCCTTGTCACAGATGGGTGC
TGGTCCTCAGGCGTTGGGGACACTGCTGGGTCGATGGGGTCGGATTCTGCCAGTTTCTGCTCTGCAGCCAAAG
ATGGTCAGAAGCATTGTCACTTCAGTAACATCAAGTGCTCAAAGACATGGCAACCGTTCAGTGGTACTTAAGT
ATTCAAAATATACAACTACAGATTCTCTGACAGAAACCAGCACGGGGTCTTCACCTTCATTCACCCCACAGGC
GACATGCGAGGGAGAACAGCATCTCAGTGGTGATTTCCAAACCAAGCCTTTGTTTTCGGTGTGGGGTTTTGGG
GGTTTGCTTTAATGTTTTTGAAATTGTAAATGTTGGGCTTTTTATTTTGATGTAAACTGAGAATAATGGCATT
TTAGGGCCTGTGACCAAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTCT
GGGAGGGGTTAA (up primer)

LOCUS BC064947 3597 bp mRNA linear PRI 21-JUL-2006
DEFINITION **Homo sapiens sphingomyelin phosphodiesterase 4, neutral membrane (neutral sphingomyelinase-3), mRNA (cDNA clone MGC:70674 IMAGE:6023304), complete cds.**
ACCESSION BC064947 VERSION BC064947.1 GI:40674044

```
   1 cggaatggcg gcttccatct ctgaggcgag cgacgctatg gatcccacag tggtttgcta
  61 agaaggccat tttcaactct ccactggagg ctgctatggc gttccctcac ctgcagcagc
 121 ccagctttct actggctagc ctgaaagctg actctataaa taagcccttt gcacagcagt
 181 gccaagactt ggttaaagtc attgaggact ttccagcaaa gtggcccaat gatgaagttg
 241 gtttataagc ttcaagctga agactataag ttcgactttc ctgtctccta cttgcctggt
  //
2341 cccctcccat gcacgctgct gctcaccctg ggctatgtcc tctacgcctc tgccatgaca
2401 ctgctgaccg agcgggggaa gctgcaccag ccctgaaggt gtcagctgcc ttcagagcag
2461 gctggaggga tttgccacac agccccaccc ttgggctgag aggacctggg aagcccctcc
  //
2941 cagagagtgg ctgtcctgat tcttcactgt gaggggcgtt cttcatgttc tcccagctgt
3001 tccaagactg ggccgtagaa ttccatgttt caggagccta agaccctccc agagcccagg
3061 ggcttcaccg cagaccccaa gccattgagc acatcaccca aagcagtggc caacatcgcg
```

```
  //
3421 ttttcggtgt ggggttttgg gggtttgctt taatgttttt gaaattgtaa atgttgggct
3481 ttttattttg atgtaaactg agaataatgg cattttaggg cctgtgacca aaaatgaagc
3541 ttgtaacgac catggatctg aataaacatg tccttgcttc tgaaaaaaaa aaaaaaa //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame):
As CDS and phage insert express different proteins, whats the point?

```
Frame 1: MFQEPKTLPEPRGFTADPKPLSTSPKAVANIADPCALSQMGAGPQALGTLLGRWG
         RILPVSALQPKMVRSIVTSVTSSAQRHGNRSVVLKYSKYTTTDSLTETSTGSSPS
         FTPQATCEGEQHLSGDFQTKPLFSVWGFGGLL
Frame 2: CFRSLRPSQSPGASPQTPSH
Frame 3: VSGA
```

8.3.3.3.4 Human liver tumour

Liver Tumour cDNA Library – Round 9 – Plaque C1

GGTTCTGGGTAC**GAATTCA**AGCGTGGATACCGTCAAGCTGGCCAAGCACAAAGAGCTTGCTCCCTACGATGAG
AACTGGTTCTACACGCGAGCTGCTTCCACAGCGCGGCACCTGTACCTCCGGGGTGGCGCTGGGGTTGGCTCCA
TGACCAAGATCTATGGGGGACGTCAGAGAAACGGCGTCATGCCCAGCCACTTCAGCCGAGGCTCCAAGAGTGT
GGCCCGCCGGGTCCTCCAAGCCCTGGAGGGGCTGAAAATGGTGGAAAAGGACCAAGATGGCGGCCGCAAACTG
ACACCTCAGGGACAAAGAGATCTGGACAGAATCGCCGGACAGGTGGCAGCTGCCAACAAGAAGCATTAGAACA
AACCATGCTGGGTTAATAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTC
TTGAGGGGTACATAAGTTACTCGAGTGCGGGGGGGCTTATTAACCCAGCAAGGTTTGTTCTAATG (up
primer)

LOCUS NM_001022 872 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens ribosomal protein S19 (RPS19), mRNA.** ACCESSION
NM_001022 VERSION NM_001022.3 GI:48255921

```
  1 gtactttcgc catcatagta ttctccacca ctgttccttc cagccacgaa cgacgcaaac
 61 gaagccaagt tcccccagct ccgaacagga gctctctatc ctctctctat tacactccgg
121 gagaaggaaa cgcgggagga aacccaggcc tccacgcgcg acccccttggc cctcccccttt
181 acctctccac ccctcactag acaccctccc ctctaggcgg ggacgaactt tcgccctgag
241 agaggcggag cctcagcgtc taccctcgct ctcgcgagct ttcggaactc tcgcgagacc
301 ctacgcccga cttgtgcgcc cgggaaaccc cgtcgttccc tttcccctgg ctggcagcgc
361 ggaggccgca cg
```
atgcctgg agttactgta aaagacgtga accagcagga gttcgtcaga
gctctggcag ccttcctcaa aaagtccggg aagctgaaag tccccgaatg ggtggatacc
gtcaagctgg ccaagcacaa agagcttgct ccctacgatg agaactggtt ctacacgcga
gctgcttcca cagcgcggca cctgtacctc cggggtggcg ctggggttgg ctccatgacc
aagatctatg ggggacgtca gagaaacggc gtcatgccca gccacttcag ccgaggctcc
aagagtgtgg cccgccgggt cctccaagcc ctggaggggc tgaaaatggt ggaaaaggac
caagatggcg gccgcaaact gacacctcag ggacaaagag atctggacag aatcgccgga
caggtggcag ctgccaacaa gaagcattag aacaaaccat gctgggttaa taaattgcct
841 cattcgtaaa aaaaaaaaaa aaaaaaaaaa aa //

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SVDTVKLAKHKELAPYDENWFYTRAASTARHLYLRGGAGVGSMTKIYGGRQRNGVMP**
SHFSRGSKSVARRVLQALEGLKMVEKDQDGGRKLTPQGQRDLDRIAGQVAAANKKH
Frame 2: AWIPSSWPSTKSLLPTMRTGSTRELLPQRGTCTSGVALGLAP
Frame 3: RGYRQAGQAQRACSLR

Liver Tumour cDNA Library – Round 9 – Plaque E1

CCC**GAATTCA**GCCAAGCACAAAGAGCTTGCTCCCTACGATGAGAACTGGTTCTACACGCGAGCTGCTTCCACA
GCGCGGCACCTGTACCTCCGGGGTGGCGCTGGGGTTGGCTCCATGACCAAGATCTATGGGGGACGTCAGAGAA
ACGGCGTCATGCCCAGCCACTTCAGCCGAGGCTCCAAGAGTGTGGCCCGCCGGGTCCTCCAAGCCCTGGAGGG
GCTGAAAATGGTGGAAAAGGACCAAGATGGCGGCCGCAAACTGACACCTCAGGGACAAAGAGATCTGGACAGA
ATCGCCGGACAGGTGGCAGCTGCCAACAAGAAGCATTAGAACAAACCATGCTGGGTTAATAAATTGCCTCATT
CGTAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTCTTGGAGGGGTTA
(up primer)

LOCUS NM_001022 872 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens ribosomal protein S19 (RPS19), mRNA.** ACCESSION
NM_001022 VERSION NM_001022.3 GI:48255921

```
  1 gtactttcgc catcatagta ttctccacca ctgttccttc agccacgaa cgacgcaaac
 61 gaagccaagt tcccccagct ccgaacagga gctctctatc ctctctctat tacactccgg
121 gagaaggaaa cgcgggagga aacccaggcc tccacgcgcg acccccttggc cctcccccttt
181 acctctccac ccctcactag acaccctccc ctctaggcgg ggacgaactt tcgccctgag
241 agaggcggag cctcagcgtc taccctcgct ctcgcgagct ttcggaactc tcgcgagacc
301 ctacgcccga cttgtgcgcc cgggaaaccc cgtcgttccc tttcccctgg ctggcagcgc
361 ggaggccgca cgatgcctgg agttactgta aaagacgtga accagcagga gttcgtcaga
421 gctctggcag ccttcctcaa aaagtccggg aagctgaaag tccccgaatg ggtggatacc
481 gtcaagctgg ccaagcacaa agagcttgct ccctacgatg agaactggtt ctacacgcga
541 gctgcttcca cagcgcggca cctgtacctc cggggtggcg ctggggttgg ctccatgacc
601 aagatctatg ggggacgtca gagaaacggc gtcatgccca gccacttcag ccgaggctcc
661 aagagtgtgg cccgccgggt cctccaagcc ctggaggggc tgaaaatggt ggaaaaggac
721 caagatggcg gccgcaaact gacacctcag ggacaaagag atctggacag aatcgccgga
781 caggtggcag ctgccaacaa gaagcattag aacaaaccat gctgggttaa taaattgcct
841 cattcgtaaa aaaaaaaaaa aaaaaaaaaa aa //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SAKHKELAPYDENWFYTRAASTARHLYLRGGAGVGSMTKIYGGRQRNGVMPSHFSR**
GSKSVARRVLQALEGLKMVEKDQDGGRKLTPQGQRDLDRIAGQVAAANKKH
Frame 2: QPSTKSLLPTMRTGSTRELLPQRGTCTSGVALGLAP
Frame 3: SQAQRACSLR

Liver Tumour cDNA Library – Round 9 – Plaque D1

ACTGTGGGTAC**GAATTCA**AGCGATGGGTGGATACCGTCAAGCTGGCCAAGCACAAAGAGCTTGCTCCCTACGA
TGAGAACTGGTTCTACACGCGAGCTGCTTCCACAGCGCGGCACCTGTACCTCCGGGGTGGCGCTGGGGTTGGC
TCCATGACCAAGATCTATGGGGGACGTCAGAGAAACGGCGTCATGCCCAGCCACTTCAGCCGAGGCTCCAAGA
GTGTGGCCCGCCGGGTCCTCCAAGCCCTGGAGGGGCTGAAAATGGTGGAAAAGGACCAAGATGGCGGCCGCAA
ACTGACACCTCAGGGACAAAGAGATCTGGACAGAATCGCCGGACAGGTGGCAGCTGCCAACAAGAAGCATTAG
AACAAACCATGCTGGGTTAATAAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAAC
GGGTCTTGGAGGGGTTAA (up primer)

LOCUS NM_001022 872 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens ribosomal protein S19 (RPS19), mRNA.** ACCESSION
NM_001022 VERSION NM_001022.3 GI:48255921

```
  1 gtactttcgc catcatagta ttctccacca ctgttccttc agccacgaa cgacgcaaac
 //
301 ctacgcccga cttgtgcgcc cgggaaaccc cgtcgttccc tttcccctgg ctggcagcgc
361 ggaggccgca cgatgcctgg agttactgta aaagacgtga accagcagga gttcgtcaga
```

421 **gctctggcag ccttcctcaa aaagtccggg aagctgaaag tccccgaatg ggtggatacc**
481 **gtcaagctgg ccaagcacaa agagcttgct ccctacgatg agaactggtt ctacacgcga**
//
721 **caagatggcg gccgcaaact gacacctcag ggacaaagag atctggacag aatcgccgga**
781 **caggtggcag ctgccaacaa gaagcattag** aacaaaccat gctgggttaa taaattgcct
841 cattcgtaaa aaaaaaaaaa aaaaaaaaaa aa //

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame):

Frame 1: **SEWVDTVKLAKHKELAPYDENWFYTRAASTARHLYLRGGAGVGSMTKIYGGRQRNGVM**
PSHFSRGSKSVARRVLQALEGLKMVEKDQDGGRKLTPQGQRDLDRIAGQVAAANKKH
Frame 2: ANGWIPSSWPSTKSLLPTMRTGSTRELLPQRGTCTSGVALGLAP
Frame 3: RMGGYRQAGQAQRACSLR

Liver Tumour cDNA Library – Round 9 – Plaque E1

CCC**GAATTCG**AGCCAAGCACAAAGAGCTTGCTCCCTACGATGAGAACTGGTTCTACACGCGAGCTGCTTCCAC
AGCGCGGCACCTGTACCTCCGGGGTGGCGCTGGGGTTGGCTCCATGACCAAGATCTATGGGGGACGTCAGAGA
AACGGCGTCATGCCCAGCCACTTCAGCCGAGGCTCCAAGAGTGTGGCCCGCCGGGTCCTCCAAGCCCTGGAGG
GGCTGAAAATGGTGGAAAAGGACCAAGATGGCGGCCGCAAACTGACACCTCAGGGACAAAGAGATCTGGACAG
AATCGCCGGACAGGTGGCAGCTGCCAACAAGAAGCATTAGAACAAACCATGCTGGGTTAATAAATTGCCTCAT
TCGTAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTCTTGGAGGGGTTA
(up primer)

Best out of dozens of hits!!! Longest match is plus/minus, many others show identity of 80-90%, but not more!

LOCUS NM_001040446 5187 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens myotubularin related protein 12 (MTMR12), mRNA.**
ACCESSION NM_001040446 VERSION NM_001040446.1 GI:94721262

No hit, protein sequence also doesn't result in conlusive hit.

Frame 1: QFEGFGLGKNKAGTNSNTI
Frame 2: NLRALVWGKTKPAQIQIQFRT
Frame 3: I

Liver Tumour cDNA Library – Round 9 – Plaque C2

CTGGGAATC**GAATTCA**AGCGTGGATACCGTCAAGCTGGCCAAGCACAAAGAGCTTGCTCCCTACGATGAGAAC
TGGTTCTACACGCGAGCTGCTTCCACAGCGCGGCACCTGTACCTCCGGGGTGGCGCTGGGGTTGGCTCCATGA
CCAAGATCTATGGGGGACGTCAGAGAAACGGCGTCATGCCCAGCCACTTCAGCCGAGGCTCCAAGAGTGTGGC
CCGCCGGGTCCTCCAAGCCCTGGAGGGGCTGAAAATGGTGGAAAAGGACCAAGATGGCGGCCGCAAACTGACA
CCTCAGGGACAAAGAGATCTGGACAGAATCGCCGGACAGGTGGCAGCTGCCAACAAGAAGCATTAGAACAAAC
CATGCTGGGTTAATAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTCTTG
AGGGGTTAA (up primer)

LOCUS NM_001022 872 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens ribosomal protein S19 (RPS19), mRNA.** ACCESSION
NM_001022 VERSION NM_001022.3 GI:48255921

1 gtactttcgc catcatagta ttctccacca ctgttccttc cagccacgaa cgacgcaaac
//

301 ctacgcccga cttgtgcgcc cgggaaaccc cgtcgttccc tttcccctgg ctggcagcgc
361 ggaggccgca **cgatgcctgg agttactgta aaagacgtga accagcagga gttcgtcaga**
421 **gctctggcag ccttcctcaa aaagtccggg aagctgaaag tccccgaatg ggtggatacc**
481 **gtcaagctgg ccaagcacaa agagcttgct ccctacgatg agaactggtt ctacacgcga**
//
721 **caagatggcg gccgcaaact gacacctcag ggacaaagag atctggacag aatcgccgga**
781 **caggtggcag ctgccaacaa gaagcattag** aacaaaccat gctgggttaa taaattgcct
841 cattcgtaaa aaaaaaaaaa aaaaaaaaaa aa //

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SVDTVKLAKHKELAPYDENWFYTRAASTARHLYLRGGAGVGSMTKIYGGRQRNGVMPS**
HFSRGSKSVARRVLQALEGLKMVEKDQDGGRKLTPQGQRDLDRIAGQVAAANKKH
Frame 2: AWIPSSWPSTKSLLPTMRTGSTRELLPQRGTCTSGVALGLAP
Frame 3: RGYRQAGQAQRACSLR

Liver Tumour cDNA Library – Round 9 – Plaque B2

AAGGGAGGGTGGGCCACCGCTGGACACTCTTGGGCCGAGGGGCAAGGGTGGACATAGCATTGGCATGGCAGCA
AGAGACTTGGCTGAGGGCAGGGGTGGAAGGATGTCTACTGCTGAAGACCAAATCCCTTGCCACTCTGTCGGTT
GCTACTAAGTCAAACGCAAACAGAAACAGGAAAAGACAGCAGGAGCACAGAGGCCAGCATCTCCCTTTCCTTA
TGCCATGGAGCCAGGTGCCAGGCGTGGGCTGGTCCATCTCCCTCCTCAGCCTCTGCCAAAGGGCTGCTCCAGT
GCTCCCCTCAGTGGAACTCCTCAGCCAGGCGCTCTGAAACCAACCACAGTCCACTGAAGTTGACCACCCTACC
GAGACAGACCAAGGCTCACAGGTGAAAACCCCTTTCAACAAAATTCCTTCCAGGTAGGAATCAGAATATGATG
CGTTTCCCATCTGCCGGGTTCTGCTCCATGGGACAGTAGGGACACTTCAGCTTTCCTCCATTAATGAGCTTAT
TGAGTGCATCTCGGGAGATAACATGGCCACAGATGAGCTTGATGGGAGGGTTGGAATCTGACGTCTGCTGGCG
GAGGATGGGGCAAGCGAACACGGAGTGGTACCAGCACTTCATGCCTAGTTCAATCTCAAGCTTGCGGCCGCAC
TCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTTTTGAGGGGTTACAAACTAGGTTACTCGAGTGCG
GCGCAAGCTTGAGATTGAACTAGGCCTGAAGTGCTGGTACACTCGTGTTCGCTTGCCCCATTCTCCGCCGCAG
ACGTCAGATTCCACCCCTC (up primer)

LOCUS NM_022762 1825 bp mRNA linear PRI 15-AUG-2011
DEFINITION **Homo sapiens required for meiotic nuclear division 5 homolog B (S. cerevisiae) (RMND5B), mRNA.**
ACCESSION NM_022762 VERSION NM_022762.3 GI:34147687

Gene in backwards

Liver Tumour cDNA Library – Round 9 – Plaque A3

ACCCCGGCACAAGA**GAATTCC**CAATTTGAGGGCTTTGGTTTGGGGAAAAACAAAGCCGGCACAAATTCAAATA
CAATTTAGAACATGACTTCAAGGCTGGGCGCAATGGCTCACACCCGCCCATAATCCCAGCACTTTGGGAAGTC
AAGGTGGGCGGATTGCTTGAGTCCAGGAGTTCGAGACCAGCCTGGGCAACATGGAAGAAACCCCTGTCTCTAC
TAAAAATACAAAAAATTAGCCAGGTGTGGTGGTGTGCACCTGTAGTCCCAGCTACTCGGGAGGCTGAGGTGGG
AAAATGGCTTGAGCCTAGGAGATGGAGGCTGCAGTGAGCTGTGATCGTGCCACCACACTCCAGCTTAGGCGAC
AGAGCCAGACCGTCTTAAAAAACAAAACAAAAAATAATATGACTTCAAAATGCAGGTGGGGGTGTGAGCAGTG
CACCTGGGCTCTTGCTGTCTGGAGCTTTGTTAGTCTAATTTCAAAGCTTTAAGCAACACACACACACACACAC
GCCTCATGTAATGGACTTTTATTACAAAGAAAAAAATTCCAGTGAATTGTGCAGAAATGCTGGTTTTTACACC
ATCCTAAAGAAAAACTTTACAAGGGTGTTTTGGAGTAGAAAAAAGGTTATAAAGTTGGAATCTTAAATTGTAA
AATTAACCATTGAGTGTCAAAGTTCTAAAAGCAGAACTCATTTTGTGCAATGAACATAAGGAAAGACTACTGT
ATAGGTTTTTTTTTTTTCTCCTTTTAAATGAAGAAAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTT
GGGGCCTCTAAACGGGTCTTGAGGGTTAACAACTAGGTCACTCGAGTGCGGCGCAAGCTTTTCTTTCATTTAA
AGGAGAAAAAAAAAACTATACGGAGTCTTTCCATATGTTCATTGCACAAATGAGTCTGCTTTACACTTTGACT
CAGGGTATATTTTCAAT (up primer)

LOCUS NM_004776 4743 bp mRNA linear PRI 26-NOV-2011
DEFINITION **Homo sapiens UDP-Gal:betaGlcNAc beta 1,4-galactosyltransferase, polypeptide 5(B4GALT5), mRNA.**
ACCESSION NM_004776 VERSION NM_004776.3 GI:163644314

Gene in backwards

Liver Tumour cDNA Library – Round 9 – Plaque B3

AAAG**GAATTCC**AAAAAACCACATGACGATATGGATGAGGATGATAATGTATCAATGGGTGGGCCTGATAGTCC
TGATTCAGTGGATCCCGTTGAACCAATGCCAACCATGACTGATCAAACAACACTTGTTCCAAATGAGGAAGAA
GCATTTGCATTGGAGCCTATTGATATAACTGTTAAAGAAACAAAAGCCAAGAGGAAGAGGAAGCTAATTGTTG
ACAGTGTCAAAGAGTTGGATAGCAAGACAATTAGAGCCCAACTTAGTGATTATTCAGATATTGTTACTACTTT
GGATCTGGCACCGCCCACCAAGAAATTGATGATGTGGAAAGAGACAGGAGGAGTAGAAAAACTGTTTTCTTTA
CCTGCTCAGCCTTTGTGGAATAACAGACTACTGAAGCTCTTTACACGCTGTCTTACACCGCTTGTACCAGAAG
ACCTTAGAAAAAGGAGGAAAGGAGGAGAGGCAGATAATTTGGATGAATTCCTCAAAGAATTTGAAAATCCAGA
GGTTCCTAGAGAGGACCAGCAACAGCAGCATCAGCAGCGTGATGTTATCGGTTTGTTCTGTTGGAATAGTTCC
TTGATGAAGTTCACATGAGTATTTGGCTTAAAATTTACCTTTATACTGTCAGAGTGAAAGAAATGAGGAATCA
AAGATATACTCAAATGTTTGGGCAGGGAAATGCATTGGATCTGATTGCCTTATTTAAGTATGAAGCCTCCTGT
CTTTCCAGAGTTTTAAGGTTCAGAACTGGCAGGGTAGATAGTCTGCTAAAGAATTGTCCTACATTGTACTGAG
GGTAGGACTTGTCCTTTTAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGC
TTTGGAAGGGGGTTACA (up primer)

LOCUS NM_006265 3773 bp mRNA linear PRI 13-AUG-2011
DEFINITION **Homo sapiens RAD21 homolog (S. pombe) (RAD21), mRNA.**
ACCESSION NM_006265 VERSION NM_006265.2 GI:208879448

```
   1 aggcgcacag gtaccatttt gaccgtaaac atcctgccga tttgaaccga ggatttgggc
  61 ggcaggaaga gccgcggcgt aacggcagcc atcttgtttg tttgagtgaa tcggaaagga
 121 ggcgccggct gtggcggcgg cgggagctgc tcggaagcta cacctcgcaa gggctccccc
 181 ctttccccac cccctccccc gacccttttc ccctccccgg gccacccagc ccgcccaact
 241 cccagcggag agcaaggttt tcttctgttt tcatagccag ccagaacaat gttctacgca
 301 catttgttc tcagtaaaag agggcctctg gccaaaattt ggctagcggc ccattgggat
 361 aagaagctaa ccaaagccca tgtgttcgag tgtaatttag agagcagcgt ggagagtatc
 //
 961 gatggtggaa tattagatga caaacttatt agtaataatg atggcggtat ctttgatgat
1021 ccccctgccc tctctgaggc aggggtgatg ttgccagagc agcctgcaca tgacgatatg
1081 gatgaggatg ataatgtatc aatgggtggg cctgatagtc ctgattcagt ggatcccgtt
1141 gaaccaatgc caaccatgac tgatcaaaca acacttgttc caaatgagga agaagcattt
//
1501 aggaaaggag gagaggcaga taatttggat gaattcctca aagaatttga aaatccagag
1561 gttcctagag aggaccagca acagcagcat cagcagcgtg atgttatcga tgagcccatt
1621 attgaagagc caagccgcct ccaggagtca gtgatggagg ccagcagaac aaacatagat
1681 gagtcagcta tgcctccacc accacctcag ggagttaagc gaaaagctgg acaaattgac
//
2041 tgtcgaaata cgaacagaaa acaagctgcc gcaaagttct acagcttctt ggttcttaaa
2101 aagcagcaag ctattgagct gacacaggaa gaaccgtaca gtgacatcat cgcaacacct
2161 ggaccaaggt tccatattat ataaggagct agaagcatta tagctagtgt ttgattcact
2221 agtgcttaca aattgccccc atgtgtaggg gacacagaac cctttgagaa aacttagatt
//
3601 gaatgtttga ttttagaaag gtcattagtt tcttgttaca cattttgtta gtctggtttt
3661 tgttgcttat cgggtttaat attgttcttg aaaatagttg atgctatgtt atgtataact
3721 tttctaataa aagttgtgtt ataagctgta aaaaaaaaaa aaaaaaaaaa aaa //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **KKPHDDMDEDDNVSMGGPDSPDSVDPVEPMPTMTDQTTLVPNEEEAFALEPIDI TVKETKAKRKRKLIVDSVKELDSKTIRAQLSDYSDIVTTLDLAPPTKKLMMWKE TGGVEKLFSLPAQPLWNNRLLKLFTRCLTPLVPEDLRKRRKGGEADNLDEFLKE FENPEVPREDQQQQHQQRDVIGLFCWNSSLMKFT**
(Nucleotide sequence from phage is 889bp long, but match with above gene only up to position 558)
Frame 2: KNHMTIWMRMIMYQWVGLIVLIQWIPLNQCQP
Frame 3: KTT

Liver Tumour cDNA Library – Round 9 – Plaque H3

GGTTCTGGGTAC**GAATTCA**AGCGTGGAACCGTCAAGCTGGCCAAGCACAAAGAGCTTGCTCCCTACGATGAGA
ACTGGTTCTACACGCGAGCTGCTTCCACAGCGCGGCACCTGTACCTCCGGGGTGGCGCTGGGGTTGGCTCCAT
GACCAAGATCTATGGGGGACGTCAGAGAAACGGCGTCATGCCCAGCCACTTCAGCCGAGGCTCCAAGAGTGTG
GCCCGCCGGGTCCTCCAAGCCCTGGAGGGGCTGAAAATGGTGGAAAAGGACCAAGATGGCGGCCGCAAACTGA
CACCTCAGGGACAAAGAGATCTGGACAGAATCGCCGGACAGGTGGCAGCTGCCAACAAGAAGCATTAGAACAA
ACCATGCTGGGTTAATAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTCT
TGAGGGGTACATAAGTTACTCGAGTGCGGGGGGGCTTATTAACCCAGCAAGGTTTGTTCTAATG (up primer)

LOCUS NM_001022 872 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens ribosomal protein S19 (RPS19), mRNA.** ACCESSION NM_001022 VERSION NM_001022.3 GI:48255921

```
  1 gtactttcgc catcatagta ttctccacca ctgttccttc cagccacgaa cgacgcaaac
 //
301 ctacgcccga cttgtgcgcc cgggaaaccc cgtcgttccc tttcccctgg ctggcagcgc
361 ggaggccgca cg atgcctgg agttactgta aaagacgtga accagcagga gttcgtcaga
421 gctctggcag ccttcctcaa aaagtccggg aagctgaaag tccccgaatg ggtggatacc
481 gtcaagctgg ccaagcacaa agagcttgct ccctacgatg agaactggtt ctacacgcga
 //
721 caagatggcg gccgcaaact gacacctcag ggacaaagag atctggacag aatcgccgga
781 caggtggcag ctgccaacaa gaagcattag aacaaaccat gctgggttaa taaattgcct
841 cattcgtaaa aaaaaaaaaa aaaaaaaaaa aa //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SVDTVKLAKHKELAPYDENWFYTRAASTARHLYLRGGAGVGSMTKIYGGRQRNGVMPS HFSRGSKSVARRVLQALEGLKMVEKDQDGGRKLTPQGQRDLDRIAGQVAAANKKH**
Frame 2: AWIPSSWPSTKSLLPTMRTGSTRELLPQRGTCTSGVALGLAP
Frame 3: RGYRQAGQAQRACSLR

8.3.3.3.5 Human Lung Tumour

Lung Tumour cDNA Library – Round 9 – Plaque A4

CGCCTTTCCGGTGAGAC**GAATTCA**AGCCGAGGCTACGTGAAGGACAGTTTGCCTGGAGACATTTCCACTGGTA
CCTTACCAATGAGGGTATCCAGTATCTCCGTGATTACCTTCATCTGCCCCCGGAGATTGTGCCTGCCACCCTA
CGCCGTAGCCGTCCAGAGACTGGCAGGCCTCGGCCTAAAGGTCTGGAGGGTGAGCGACCTGCGAGACTCACAA
GAGGGGAAGCTGACAGAGATACCTACAGACGGAGTGCTGTGCCACCTGGTGCCGACAAGAAAGCCGAGGCTGG
GGCTGGGTCAGCAACCGAATTCCAGTTTAGAGGCGGATTTGGTCGTGGACGTGGTCAGCCTCCTCAGTAAAAT
TGGAGAGGATTCTTTTGCATTGAATAAACCTCTTTCCCAAAAGACTGTCAGAGGCGTGAGTGCTGCAAAAGAA
CAACAACACCAAAACGCTGGTCCTGCTCATTTTCTACAACTGTGTACACACACACACTCACTTGCATCCAATA
AAAACAATCGTAGCAGATTTGCTTATTTAATGAACTTTATCCACATTATCTCTTCAAATTGCCAGGAGGCTGG
AAAAATTAAGGGTATTCACCATGCCCAAAGTCCACCGGTATTCAGGCTTCAAAATTTTTTTTCTTTTTTTTTT
TTTCATTTTTTGTTTCCAATGGGATTGGTGAAGACACCTGCATTTTTAGTGCCAGGAAGATAAAAAGCTTGCG

GCCCGCCCCTCGAGTAACTAGTTTAACCCCCTTTGGGCCTCTAAACGGGTCTTTGGAGGGGTTTA (up primer)

LOCUS NM_001204091 666 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens ribosomal protein S10 (RPS10), transcript variant 3, mRNA.**
ACCESSION NM_001204091 VERSION NM_001204091.1 GI:323276699

```
  1 ggcggggggc gggtccacgc cagcccggaa gagacgcagc accgcgcatg ctccttcctt
 61 tccagccccg gtaccggacc ctgcagccgc agaggtgaat gttgatgcct aagaagaacc
121 ggattgccat ttatgaactc ctttttaagg agggagtcat ggtggccaag aaggatgtcc
181 acatgcctaa gcacccggag ctggcagaca agaatgtgcc caaccttcat gtcatgaagg
241 ccatgcagtc tctcaagtcc cgaggctacg tgaaggaaca gtttgcctgg agacatttct
301 actggtacct taccaatgag ggtatccagt atctccgtga ttaccttcat ctgcccccgg
361 agattgtgcc tgccacccta cgccgtagcc gtccagagac tggcaggcct cggcctaaag
421 gtctggaggg tgagcgacct gcgagactca caagagggga agctgacaga gatacctaca
481 gacggagtgc tgtgccacct ggtgccgaca agaaagccga ggctggggct gggtcagcaa
541 ccgaattcca gtttagaggc ggatttggtc gtggacgtgg tcagccacct cagtaaaatt
601 ggagaggatt cttttgcatt gaataaactt acagccaaaa aaccttaaaa aaaaaaaaaa
661 aaaaaa //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SRGYVKEQFAWRHFHWYLTNEGIQYLRDYLHLPPEIVPATLRRSRPETGRPRPKGL EGERPARLTRGEADRDTYRRSAVPPGADKKAEAGAGSATEFQFRGGFGRGRGQPPQ**
Frame 2: AEAT
Frame 3: PRLREGTVCLETFPLVPYQ

Lung Tumour cDNA Library – Round 9 – Plaque C4

AACTGAACCAGGCTAGAC**GAATTCA**AGCGTCAAGCTGGCCAGCACAAAGAGCTTGCTCCCTACGATGAGAACT
GGTTCTACACGCGAGCTGCTTCCACAGCGCGGCACCTGTACCTCCGGGGTGGCGCTGGGGTTGGCTCCATGAC
CAAGATCTATGGGGGACGTCAGAGAAACGGCGTCATGCCCAGCCACTTCAGCCGAGGCTCCAAGAGTGTGGCC
CGCCGGGTCCTCCAAGCCCTGGAGGGGCTGAAAATGGTGGAAAAGGACCAAGATGGCGGCCGCAAACTGACAC
CTCAGGGACAAAGAGATCTGGACAGAATCGCCGGACAGGTGGCAGCTGCCAACAAGAAGCATTAGAACAAACC
ATGCTGGGTTAATAAATTGCCTCATTCGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGTTTGGGGCCCCCCC
CCAAAAAATTTTTAACCCCTTGGGGCCCTCAAAGGGGTTTGGGGGGGGTATAA (up primer)

LOCUS NM_001022 872 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens ribosomal protein S19 (RPS19), mRNA.** ACCESSION NM_001022 VERSION NM_001022.3 GI:48255921

```
  1 gtactttcgc catcatagta ttctccacca ctgttccttc cagccacgaa cgacgcaaac
 //
301 ctacgcccga cttgtgcgcc cgggaaaccc cgtcgttccc tttcccctgg ctggcagcgc
361 ggaggccgca cgatgcctgg agttactgta aaagacgtga accagcagga gttcgtcaga
421 gctctggcag ccttcctcaa aaagtccggg aagctgaaag tccccgaatg ggtggatacc
481 gtcaagctgg ccaagcacaa agagcttgct ccctacgatg agaactggtt ctacacgcga
 //
721 caagatggcg gccgcaaact gacacctcag ggacaaagag atctggacag aatcgccgga
781 caggtggcag ctgccaacaa gaagcattag aacaaaccat gctgggttaa taaattgcct
841 cattcgtaaa aaaaaaaaaa aaaaaaaaaa aa //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SVKLAKHKELAPYDENWFYTRAASTARHLYLRGGAGVGSMTKIYGGRQRNGVMPSH FSRGSKSVARRVLQALEGLKMVEKDQDGGRKLTPQGQRDLDRIAGQVAAANKKH**
Frame 2: ASSWPSTKSLLPTMRTGSTRELLPQRGTCTSGVALGLAP
Frame 3: RQAGQAQRACSLR

Lung Tumour cDNA Library – Round 9 – Plaque C5

GCCCGGGGCAC**GAATTCA**AGCGATGTCCACATGCCTAAGCACCCGGAGCTGGCAGACAAGAATGTGCCCAACC
TTCATGTCATGAAGGCCATGCAGTCTCTCAAGTCCCGAGGCTACGTGAAGGAACAGTTTGCCTGGAGACATTT
CTACTGGTACCTTACCAATGAGGGTATCCAGTATCTCCGTGATTACCTTCATCTGCCCCCGGAGATTGTGCCT
GCCACCCTACGCCGTAGCCGTCCAGAGACTGGCAGGCCTCGGCCTAAAGGTCTGGAGGGTGAGCGACCTGCGA
GACTCACAAGAGGGGAAGCTGACAGAGATACCTACAGACGGAGTGCTGTGCCACCTGGTGCCGACAAGAAAGC
CGAGGCTGGGGCTGGGTCAGCAACCGAATTCCAGTTTAGAGGCGGATTTGGTCGTGGACGTGGTCAGCCACCT
CAGTAAAATTGGAGAGGATTCTTTTGCATTGAATAAACTTACAGCCAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAGTTGGGGGCCCCCCCCAAAAAAATTTTAACCCCCTGGGGGCCCTAAAAGGGGTTTGGGGGGGGGATAAA
(up primer)

Matches variants 1 – 3 & pseudogene 7
LOCUS NM_001204091 666 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens ribosomal protein S10 (RPS10), transcript variant 3, mRNA.**
ACCESSION NM_001204091 VERSION NM_001204091.1 GI:323276699

```
  1 ggcggggggc gggtccacgc cagcccggaa gagacgcagc accgcgcatg ctccttcctt
 61 tccagccccg gtaccggacc ctgcagccgc agaggtgaat gttgatgcct aagaagaacc
121 ggattgccat ttatgaactc ctttttaagg agggagtcat ggtggccaag aaggatgtcc
181 acatgcctaa gcacccggag ctggcagaca agaatgtgcc caaccttcat gtcatgaagg
 //
481 gacggagtgc tgtgccacct ggtgccgaca agaaagccga ggctggggct gggtcagcaa
541 ccgaattcca gtttagaggc ggatttggtc gtggacgtgg tcagccacct cagtaaaatt
601 ggagaggatt cttttgcatt gaataaactt acagccaaaa aaccttaaaa aaaaaaaaaa
661 aaaaaa //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SDVHMPKHPELADKNVPNLHVMKAMQSLKSRGYVKEQFAWRHFYWYLTNEGIQYLR DYLHLPPEIVPATLRRSRPETGRPRPKGLEGERPARLTRGEADRDTYRRSAVPPGA DKKAEAGAGSATEFQFRGGFGRGRGQPPQ**
Frame 2: AMSTCLSTRSWQTRMCPTFMS
Frame 3: RCPHA

8.3.3.3.6 Human breast tumour

Breast Tumour cDNA Library – Round 9 – Plaque B7

TCAATCGGGAACGGCAC**GAATTCA**GCTATCAGACCTAATCAGCCTACATATGTTTCCTATATATCAAAGAAAA
ATATATTATTCACCACTACATAGTTGGGGAAAATATCACACATTTGCTATAACTAGAGGCTTACAGTGGGTTA
CCTTATATCACTAAAAATACTATCTGAGTTTTATATAATGATTCTAAATACTAACCTACTGGCATAGTTACAC
ACATTGACTCTATCTTGGCTGATGTTGCTAAAGATTTTAGGTCTTGATTCTGTCTCTGACAGATGTAAGATTC
TGTGTCTGTTGTTTGCCATGGAGACTGAGGCTACATAATGTTAAGTTAGCCGTGTGTGTGTGTGTGTGTGTGT
GTATAGGAAGAGTTTTTTTTTCTTTTTCTTTTTTTTTTTTTTTAAAAACGGAATTTCCCTTTTGTTGCCCAGG
CTGGAGTGCAGTGGCGTGATTTTGGCTCACTGAACCCTTCCTTCTCCGGGGTTCAAGCAATTCTCCTCCCTAG

CCTCCCAAGTACTGGGATAACGGGTTTTGCACCCCCCCCGTTTAGTTTTCATCTTGGAAAAAATG (up primer)

GCACGGGGTTTTCTAAGATTCTCGAAGTGCGGCCGCAGCTTTTAGGGCTTCAGAAGAGAGAAAGAAAGAGGTG
GAGGGTATTTTGCAGGTTTGCTTTCTTTGCTACTCTCCACTTTATGTTTCTTTCAATCTTCTCTAAATGTGTC
CCATTCCCTACCTCTACTCTCTGTGAAGAGGTGAAGAGAGAAGAGAAAGAAGACAGAAATGTGGACAGCATTC
AGAGTGGATTCTTCACACCGTCATGCTTCCCCACTCACCCACAGGTGAAATGTGAAAGTACAGGATTCATTAC
TAGCAATATAGCCTCCAAACTCAGAACTGTAAATACACACAATTCCCTGTAGTCATCAGGTCTCAGCAAATTC
ATGTTAGCTGCTGAGTCAATGATTAATCAAAATTAATTAGGACCCTCATACTATAGAGAACAAAACTCCTGCC
TAAGAAATTTAGCTCAGCTAACTGTGAATTATCAAGCCAAGAATTTGGAAGGGGAGAGGGCAGGAAAGGTTTT
TGTAGGACAAGTAATATCACACATATTTAGAAAGGCTCCCAGCCAGGCACGGTGGCTCACGCCTGTAATCCCA
GCACTTTGGGAGGCCGAGGCAGGCAGATTTCCTGAGGTCAGGAGTTCAAGAGCAGCCTGGCCAATGTGGTGAA
ACCCCATCTCTACTAAAAATGCAAAAACTAACTGGGCGTGGTGGCAGATGCCTGTAATCCCAGCTACTTGGGA
GGCTGAGGTAGGAGAATTGCTTGAACCCAGGAGACGAAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCC
AGCCTGGGCAACAGAGCGAGACTCCGTCTCAAAAAAAAAAAAAAAAGAAAAAGAAAAAAAAACTCTTTCTTAT
ACACACACACACACCACACACACACGGCTAACTTAACTTTATGTAGCCTCAGTCTCCGTGCAAAACAACAAGA
CACAGAATCTTACATCTGTCAGAAGACAAAATCAAGATCTAAATCCTTTTAGCAACATCAGCCAAGAATAAGA
GTCAATGGTGGTGTA (down primer)

LOCUS NM_001005862 4816 bp mRNA linear PRI 16-OCT-2011
DEFINITION **Homo sapiens v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian) (ERBB2), transcript variant 2, mRNA.**
ACCESSION NM_001005862 VERSION NM_001005862.1 GI:54792097

```
   1 gttcccggat ttttgtgggc gcctgccccg cccctcgtcc ccctgctgtg tccatatatc
  61 gaggcgatag ggttaaggga aggcggacgc ctgatgggtt aatgagcaaa ctgaagtgtt
  //
 421 cagagactca gaccctggca gccatgcctg cgcaggcagt gatgagagtg acatgtactg
 481 ttgtggacat gcacaaaagt gagtgtgcac cggcacagac atgaagctgc ggctccctgc
 541 cagtcccgag acccacctgg acatgctccg ccacctctac cagggctgcc aggtggtgca
  //
3241 gggggagcgg ctgccccagc cccccatctg caccattgat gtctacatga tcatggtcaa
3301 atgttggatg attgactctg aatgtcggcc aagattccgg gagttggtgt ctgaattctc
3361 ccgcatggcc agggaccccc agcgctttgt ggtcatccag aatgaggact tgggcccagc
3421 cagtcccttg gacagcacct tctaccgctc actgctggag gacgatgaca tgggggacct
3481 ggtggatgct gaggagtatc tggtacccca gcagggcttc ttctgtccag accctgcccc
3541 gggcgctggg ggcatggtcc accacaggca ccgcagctca tctaccagga gtggcggtgg
3601 ggacctgaca ctagggctgg agccctctga agaggaggcc cccaggtctc cactggcacc
3661 ctccgaaggg gctggctccg atgtatttga tggtgacctg ggaatggggg cagccaaggg
3721 gctgcaaagc ctccccacac atgaccccag ccctctacag cggtacagtg aggaccccac
3781 agtacccctg ccctctgaga ctgatggcta cgttgccccc ctgacctgca gcccccagcc
3841 tgaatatgtg aaccagccag atgttcggcc ccagcccccc tcgccccgag agggccctct
3901 gcctgctgcc cgacctgctg gtgccactct ggaaaggccc aagactctct ccccagggaa
3961 gaatggggtc gtcaaagacg tttttgcctt tgggggtgcc gtggagaacc ccgagtactt
4021 gacaccccag ggaggagctg cccctcagcc ccaccctcct cctgccttca gcccagcctt
4081 cgacaacctc tattactggg accaggaccc accagagcgg ggggctccac ccagcacctt
4141 caaagggaca cctacggcag agaacccaga gtacctgggt ctggacgtgc cagtgtgaac
4201 cagaaggcca agtccgcaga agccctgatg tgtcctcagg gagcagggaa ggcctgactt
4261 ctgctggcat caagaggtgg gagggccctc cgaccacttc caggggaacc tgccatgcca
  //
4741 tatggggagg caagtgtggg gggtccttct ccacacccac tttgtccatt tgcaaatata
4801 ttttggaaaa cagcta //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **NSSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQ**
GFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLAPSEGAG

SDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ
PEYVKLAAALE
Frame 2: IHPAWPGTPSALWSSRMRTWAQPVPWTAPSTAHCWRTMTWGTWWMLRSIWYPSR
ASSVQTLPRALGAWSTTGTAAHLPGVAVGT
Frame 3: FIPHGQGPPALCGHPE

Breast Tumour cDNA Library – Round 9 – Plaque D7

TAGACTTCGGTTCTGAGTC**GAATTCA**GCAGGATGTCCACATGCCTAAGCACCCGGAGCTGGCAGACAAGAATG
TGCCCAACCTTCATGTCATGAAGGCCATGCAGTCTCTCAAGTCCCGAGGCTACGTGAAGGAACAGTTTGCCTG
GAGACATTTCTACTGGTACCTTACCAATGAGGGTATCCAGTATCTCCGTGATTACCTTCATCTGCCCCCGGAG
ATTGTGCCTGCCACCCTACGCCGTAGCCGTCCAGAGACTGGCAGGCCTCGGCCTAAAGGTCTGGAGGGTGAAG
CGACTTCCAAGACTCAAAAAAGGGGAAACCTAACAAAAAACCTTACAAACGGAGTGCTGTGCCCCCTGGGGGC
CAAAAAAAACCCAAGGCGGGGGTGGGTCCACAACCAATTTCCAGTTTAAAGCGGATTTTGGCCGGGAATGGTC
CCCCCCCCCCAAAAATTTGGAAAGATCCTTTGCTTGAAAACCTTCCGCCCAAAAACTTTTAAAAAAAGAAAGG
AAAAACAAAGAGGTTGGGGGCCCCCCAAAAAAAAAAACCCCGGGGGCCCAAGGGGGGGGGGGTAA (up
primer)

GCAAGGGGTTAGTGATTCTCGAGTGCGGCCGCAAGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
TTTTTAAAGGGTTTGGGGGGGAAAATTTTTTTAAAGAAAAAAAACCCCCCCAATTTTTTTGGGGGGGGGGAAC
CCCCCCCCCAAAAAAATCCCCCCTAAAAAGGGAATTTTGTGGGGAACCCCCCCCCCCCCCCCTTTTTTTTTTG
GCCCCCGGGGGGGAAAACCCCCCCCGGGGGGGATTCCTTGCCCCCCCCCCCCCCTTGGGGAGGCCGAGGGGCCC
CCCCCCCCCAAAATTTTGGGGGGGGGCCGGCCTTCTTTTGGAGGGAAAGGGGGGGGGGGGGGGAAAAAATC
CGGGGGGGAAAAAAAGGGAATCACCAAAAAAATGGAACCCCCTTTGGGGAAGGGCCCCAAAAAAATTTCCCCC
GGAAAAATTTTTCTTCAAAAGACCCCGGGAAATAAGAAAAAGGGGGGGCTTTAAAAAATAAGGGGGGGGAAAA
TTTTTTTTCCCCCCCCCCGGGGGGTAAGGGGGGGGGAACCCTTTTTTTAAATTTGAACCCCCCCGGCCCCCCC
CGGGGGGGAAAAAACCCCCCAAAA (down primer)

LOCUS NM_001022 872 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens ribosomal protein S19 (RPS19), mRNA.**
ACCESSION NM_001022 VERSION NM_001022.3 GI:48255921

```
  1 gtactttcgc catcatagta ttctccacca ctgttccttc cagccacgaa cgacgcaaac
 61 gaagccaagt tcccccagct ccgaacagga gctctctatc ctctctctat tacactccgg
121 gagaaggaaa cgcgggagga aacccaggcc tccacgcgcg acccccttggc cctccccttt
181 acctctccac ccctcactag acaccctccc ctctaggcgg ggacgaactt tcgccctgag
241 agaggcggag cctcagcgtc taccctcgct ctcgcgagct ttcggaactc tcgcgagacc
301 ctacgcccga cttgtgcgcc cgggaaaccc cgtcgttccc tttcccctgg ctggcagcgc
361 ggaggccgca cgatgcctgg agttactgta aaagacgtga accagcagga gttcgtcaga
421 gctctggcag ccttcctcaa aaagtccggg aagctgaaag tccccgaatg ggtggatacc
481 gtcaagctgg ccaagcacaa agagcttgct ccctacgatg agaactggtt ctacacgcga
541 gctgcttcca cagcgcggca cctgtacctc cggggtggcg ctggggttgg ctccatgacc
601 aagatctatg ggggacgtca gagaaacggc gtcatgccca gccacttcag ccgaggctcc
661 aagagtgtgg cccgccgggt cctccaagcc ctggaggggc tgaaaatggt ggaaaaggac
721 caagatggcg gccgcaaact gacacctcag ggacaaagag atctggacag aatcgccgga
781 caggtggcag ctgccaacaa gaagcattag aacaaaccat gctgggttaa taaattgcct
841 cattcgtaaa aaaaaaaaaa aaaaaaaaaa aa //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **GVKLAKHKELAPYDENWFYTRAASTARHLYLRGGAGVGSMTKIYGGRQRNGVMPS**
HFSRGSKSVARRVLQALEGLKMVEKDQDGGRKLTPQGQRDLDRIAGQVAAANKKH
Frame 2: ASSWPSTKSLLPTMRTGSTRELLPQRGTCTSGVALGLAP
Frame 3: RQAGQAQRACSLR

Breast Tumour cDNA Library – Round 9 – Plaque A9

GGGCCTGGGGAC**GAATTCA**AGCGTGGAACCGTCAAGCTGGCCAAGCACAAAGAGCTTGCTCCCTACGATGAGA
ACTGGTTCTACACGCGAGCTGCTTCCACAGCGCGGCACCTGTACCTCCGGGGTGGCGCTGGGGTTGGCTCCAT
GACCAAGATCTATGGGGGACGTCAGAGAAACGGCGTCATGCCCAGCCACTTCAGCCGAGGCTCCAAGAGTGTG
GCCCGCCGGGTCCTCCAAGCCCTGGAGGGGCTGAAAATGGTGGAAAAGGACCAAGATGGCGGCCGCAAACTGA
CACCTCAGGGACAAAGAGATCTGGACAGAATCGCCGGACAGGTGGCAGCTGCCAACAAGAAGCATTAGAACAA
ACCATGCTGGGTTAATAAATTGCCTCATTCGTAAAAAAAAAAAAAAAAAAAAAAAAAAGTTGGGGGCCCCCCTC
AATTAACTATTTAACCCCTGGGGGCCTCAAAAGGGGTTTTGGGGGTTTAA (up primer)

LOCUS NM_001022 872 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens ribosomal protein S19 (RPS19), mRNA.** ACCESSION
NM_001022 VERSION NM_001022.3 GI:48255921

```
  1 gtactttcgc catcatagta ttctccacca ctgttccttc cagccacgaa cgacgcaaac
 61 gaagccaagt tcccccagct ccgaacagga gctctctatc ctctctctat tacactccgg
121 gagaaggaaa cgcgggagga aacccaggcc tccacgcgcg acccccttggc cctccccttt
181 acctctccac ccctcactag acaccctccc ctctaggcgg ggacgaactt tcgccctgag
241 agaggcggag cctcagcgtc taccctcgct ctcgcgagct ttcggaactc tcgcgagacc
301 ctacgcccga cttgtgcgcc cgggaaaccc cgtcgttccc tttcccctgg ctggcagcgc
361 ggaggccgca cgatgcctgg agttactgta aaagacgtga accagcagga gttcgtcaga
421 gctctggcag ccttcctcaa aaagtccggg aagctgaaag tccccgaatg ggtggatacc
481 gtcaagctgg ccaagcacaa agagcttgct ccctacgatg agaactggtt ctacacgcga
541 gctgcttcca cagcgcggca cctgtacctc cggggtggcg ctggggttgg ctccatgacc
601 aagatctatg ggggacgtca gagaaacggc gtcatgccca gccacttcag ccgaggctcc
661 aagagtgtgg cccgccgggt cctccaagcc ctggaggggc tgaaaatggt ggaaaaggac
721 caagatggcg gccgcaaact gacacctcag ggacaaagag atctggacag aatcgccgga
781 caggtggcag ctgccaacaa gaagcattag aacaaaccat gctgggttaa taaattgcct
841 cattcgtaaa aaaaaaaaaa aaaaaaaaaa aa //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SVDTVKLAKHKELAPYDENWFYTRAASTARHLYLRGGAGVGSMTKIYGGRQRNGVMP**
SHFSRGSKSVARRVLQALEGLKMVEKDQDGGRKLTPQGQRDLDRIAGQVAAANKKH
Frame 2: AWIPSSWPSTKSLLPTMRTGSTRELLPQRGTCTSGVALGLAP
Frame 3: RGYRQAGQAQRACSLR

Breast Tumour cDNA Library – Round 9 – Plaque D9

TCAGTCTACGATGTCCACATGCCTAAGCACCCGGAGCTGGCAGACAAGAATGTGCCCAACCTTCATGTCATGA
AGGCCATGCAGTCTCTCAAGTCCCGAGGCTACGTGAAGGAACAGTTTGCCTGGAGACATTTCTACTGGTACCT
TACCAATGAGGGTATCCAGTATCTCCGTGATTACCTTCATCTGCCCCCGGAGATTGTGCCTGCCACCCTACGC
CGTAGCCGTCCAGAGACTGGCAGGCCTCGGCCTAAAGGTCTGGAGGGTGAGCGACCTGCGAGACTCACAAGAG
GGGAAGCTGACAGAGATACCTACAGACGGAGTGCTGTGCCACCTGGTGCCGACAAGAAAGCCGAGGCTGGGGC
TGGGTCAGCAACCGAATTCCAGTTTAGAGGCGGATTTGGTCGTGGACGTGGTCAGCCACCTCAGTAAAATTGG
AGAGGATTCTTTTGCATTGAATAAACTTACAGCCAAAAAACCTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAGGGGGGGGGCCCCCCCCCAAAAAAAAAAAAACCCCCTGGGGGCCCCAAAAGG
GGGGGGGGGGGGGGTTAAA (up primer)

LOCUS NM_001204091 666 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens ribosomal protein S10 (RPS10), transcript variant 3, mRNA.**
ACCESSION NM_001204091 VERSION NM_001204091.1 GI:323276699

```
  1 ggcggggggc gggtccacgc cagcccggaa gagacgcagc accgcgcatg ctccttcctt
 61 tccagccccg gtaccggacc ctgcagccgc agaggtgaat gttgatgcct aagaagaacc
```

121 **ggattgccat ttatgaactc ctttttaagg agggagtcat ggtggccaag aaggatgtcc**
181 **acatgcctaa gcacccggag ctggcagaca agaatgtgcc caaccttcat gtcatgaagg**
241 **ccatgcagtc tctcaagtcc cgaggctacg tgaaggaaca gtttgcctgg agacatttct**
301 **actggtacct taccaatgag ggtatccagt atctccgtga ttaccttcat ctgcccccgg**
361 **agattgtgcc tgccacccta cgccgtagcc gtccagagac tggcaggcct cggcctaaag**
421 **gtctggaggg tgagcgacct gcgagactca caagagggga agctgacaga gatacctaca**
481 **gacggagtgc tgtgccacct ggtgccgaca agaaagccga ggctggggct gggtcagcaa**
541 **ccgaattcca gtttagaggc ggatttggtc gtggacgtgg tcagccacct cagtaa**aatt
601 ggagaggatt cttttgcatt gaataaactt acagccaaaa aaccttaaaa aaaaaaaaaa
661 aaaaaa //

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **KDVHMPKHPELADKNVPNLHVMKAMQSLKSRGYVKEQFAWRHFYWYLTNEGIQY**
LRDYLHLPPEIVPATLRRSRPETGRPRPKGLEGERPARLTRGEADRDTYRRSAV
VPPGADKKAEAGAGSATEFQFRGGFGRGRGQPPQ
Frame 2: AWIPSSWPSTKSLLPTMRTGSTRELLPQRGTCTSGVALGLAP
Frame 3: RGYRQAGQAQRACSLR

Breast Tumour cDNA Library – Round 9 – Plaque E9

CCCCGGGGGCAC**GAATTCA**AGCAACAAAACAAAACAAAAGCCCTATTACTCAATTATAGAGGAGAACAAACAC
AATTTAGAATGGCCAAAATATTTGAACAAACAACTAGAAAGTATACACAAATAGCCAGTGAACATTTGAAAAG
GTACTCACGTTACATATATACAAAAAGAACAAAGAACTTATATACAGAATATATAAAGAACACTTACTTTGGG
AGGCTGGGGCAGGTGGATCACTTGAGGCCAGGAGTTCAAAACAAAAGCCCTATTACTCAATTATAGAGGAGAA
CAAACACAATTTAGAATGGCCAAAATATTTGAACAAACAACTAGAAAGTATACACAAATAGCCAGTGAACATT
TGAAAATATACTCAATACCATTACTCAAGAGGAAATGTAAACCAAACCACAAGGAGCTACTGCAACACTCCCA
CTCACACAATTACATTAAGAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGG
TCTTGGAGGGGTTAA (up primer)

Multiple matches but only between bp 200 – 250!!!

Peptide encoded by Phage DNA (**Bold** = correct reading frame):

Frame 1: SNKTKQKPYYSIIEENKHNLEWPKYLNKQLESIHK
Frame 2: ATKQNKSPITQL
Frame 3: QQNKTKALLLNYRGEQTQFRMAKIFEQTTRKYTQIASEHLKRYSRYIYTKRTKNLY
TEYIKNTYFGRLGQVDHLRPGVQNKSPITQL

Breast Tumour cDNA Library – Round 9 – Plaque H9

GCTGGGGTATC**GAATTCA**AGCAAACTTTATAAAAATCTGCTGTATCTTGGACGAGACTATCCAAAAGGAGCAG
ACTATTTTAAAAAGCGTTTGAAGAACATTTTCCTTAAAAACAAAGATGTGAAGAATCCAGAGAAGATCAAAGA
ACTTATTGCACAGGGCGAATTTGTAATGAAAGAGCTAGAAGCTTTGTACTTCCTTAGGAAATACAGAGCTATG
AAACAACGCTATTATTCAGATACCAACAAAACTAATTGATCATTACTACTTTAATTTAGCTATCAGTGCCAGC
TGTTTATGTATACCAGATGTTGTAAAATAATTCTAACTTAAAATGGGAAGATATACATGTTGTGTAAAAAATC
CCTGAGCTGCCCTACTGAACTAAATAGGTTTCAACTTCTGTTCATACGGAGAAAGTATCAGCAACTTTATGCT
CAATTTTGATACAAACATAGCAATTTAGCTATACTACCGATCATAAATTAATGAGCACCCAATTTTGAATGAA
AAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTTTGGAGGGGTTAA (up primer)

LOCUS NM_001001660 1149 bp mRNA linear PRI 21-APR-2011
DEFINITION **Homo sapiens LYR motif containing 5 (LYRM5), mRNA.** ACCESSION NM_001001660 XM_370685 VERSION NM_001001660.2 GI:115430226

1 ggccggcgcc ccgccccctt tactgacagg ttgcccacct cccccaacgc caccccgctt
61 cgcagtagac ggacagagga gtcgtagcgg tcgaggcttt tgcggctccg gcgtgccgga

```
 121 aagtatgtta tgcataaaag tggataattt acatgataa tgaaaatggc caattcttta
 181 agaggagaag tactaaaact ttataaaaat ctgctgtatc ttggacgaga ctatccaaaa
 241 ggagcagact attttaaaaa gcgtttgaag aacattttcc ttaaaaacaa agatgtgaag
 301 aatccagaga agatcaaaga acttattgca cagggcgaat ttgtaatgaa agagctagaa
 361 gctttgtact tccttaggaa atacagagct atgaaacaac gctattattc agataccaac
 421 aaaactaatt gatcattact actttaattt agctatcagt gccagctgtt tatgtatacc
 481 agatgttgta aaataattct aacttaaaat gggaagatat acatgttgtg taaaaaatcc
 541 ctgagctgcc ctactgaact aaataggttt caacttctgt tcatacggag aaagtatcag
 601 caactttatg ctcaattttg atacaaacat agcaatttag ctatactacc gatcataaat
 661 taatgagcac ccaattttga atgaaaatat aatacactta atcctctaac ttaataggac
 721 ttagccaatt atttttagct tatttctcca ttttactgac aagagaatgc catttaccga
  //
1081 ttgtgtatca attaccttct ttgataaaag gaaataaata aataatgtta tcctaaaaaa
1141 aaaaaaaaa //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SKLYKNLLYLGRDYPKGADYFKKRLKNIFLKNKDVKNPEKIKELIAQGEFVMKELE ALYFLRKYRAMKQRYYSDTNKTN**
Frame 2: ANFIKICCILDETIQKEQTILKSV
Frame 3: QTL

8.3.3.3.7 Pseudomonas stutzeri

***Pseudomonas stutzeri* gDNA library – Round 9 – Plaque D10**

```
TTCAGGGCTAGAACCTCAATTAGTAAAGACTAAAGGCAGAAGGAAGAATAGATGCAGTGCCTCTATTCCCCTT
TAACTCTTACTTGTTAGGTATAAAATTGAACTTAGGGCGACAATATTTAAAATTATAAAATTTAAATGTTATT
TTCTGGCAAATCAAACAAAAGGTATAATGTAGATCAAATAAGTAGGTCACTGTTTAATAGCTCCTGCTCTGAA
AGCTTCTTACTAATTTTGATTCCTAGCACCAAGCACAATTAGCATTCAATTTATGAACTGAATATGCTAAAAC
AGAAGTCACTTTCTTATCTGAAGTCTAAAAATGATTGTACCTCCTTTAAAATGTTTATTTTGTTCAGTAATTT
AAAGAAAAAGGTCAGCAATTATGACATTAAGAACTGAAGAGTCGTGAGATCCACTCCATTAAAATGCTGAATG
ACCTGAGACCGACAATTTGGCCAGTCTTCTCTTAGCTTCACTGTTCTTTTTCACGGTCACACACAAAGAGTCA
GGTCATGCTTATTGAGGGAGAACTGAAGCAGAAATCCAAACCAAAACTGAAAAAATGAGGACACCTAAAGGCA
GCTACCAGTAAACAGTAATTCTATCTATCCTTAAGAAATCCCCTAAGCCAGGCATGGTGGCTCAAGCTTGTAA
TCCCAGCACCTTGGGAGGCCAAGGCAGAAGGATCAATCTTTGGACAATTAAGCACTTTCTGAAGTAAGTTAGA
CTGGCTAATCTTCTATATGGAAACTGTTTTTCAAAATCTTAAGTGAAAGCTTGCGGCCGCACTCGAGTAACTA
GTTAACCCCTTGGGGCCTCTAAACGGGCTTTGTAGGGGGGTTA (up primer)
```

```
LOCUS       NC_009434   4567418 bp    DNA circular BCT 16-NOV-2011
DEFINITION  Pseudomonas stutzeri A1501 chromosome, complete genome.
ACCESSION   NC_009434 VERSION     NC_009434.1  GI:146280397
```

Features in this part of subject sequence:
tRNA 2-selenouridine synthase
Score = 165 bits (182), Expect = 4e-41 Identities = 142/176 (81%), Gaps = 0/176 (0%) Strand=Plus/Plus

```
Query  7        GGCAGAAGGTCGGCACCTGCTTCAAGCAGCATGGGCAGACGGCGGCCATCGAGCTTGGCA  66
                |||||||||||||||||||| ||||||||| || ||||  || ||||||||| | |||
Sbjct  2055885  GGCAGAAGGTCGGCACCTGCTACAAGCAGCACGGCCAGAAAGCCGCCATCGAGATGGGCC  2055944
Query  67       ACCAGTTGGTCAGCGGCAGAACCAAGGACGAGCGGATTCAGGCCTGGGCCGATTTCGCCC  126
                | ||| |||||  |||||  ||||||| ||||||| | ||||||||||||||||||||||
Sbjct  2055945  ATCAGCTGGTCTCCGGCAAGACCAAGGCCGAGCGGCTGCAGGCCTGGGCCGATTTCGCCC  2056004
Query  127      GGAACAATCCGCAGGGCTACATCTACTGTTTCCGTGGCGGACTTCGTTCGCAGATC  182
                |   | | ||| | || ||| | ||||| |||||||| ||  | ||||||  |||||
Sbjct  2056005  GCGCCCACCCGGACGGTTACCTGTACTGCTTCCGTGGTGGTTTGCGTTCGAAGATC  2056060
```

tRNA 2-selenouridine synthase:

```
  1 atgaatgatg tggagcggca gaaggtcggc acctgctaca agcagcacgg ccagaaagcc
```

```
 61 gccatcgaga tgggccatca gctggtctcc ggcaagacca aggccgagcg gctgcaggcc
121 tgggccgatt tcgcccgcgc ccacccggac ggttacctgt actgcttccg tggtggtttg
181 cgttcgaaga tcgtccagca atggttgaag gacgagatgg gcatcgacta cccgcgcgtg
241 gtcggcggtt acaaggcgat gcgtcacttc ctgctcgaca ccatcgagca ggcggtggag
301 cagtgcgagt tcatactggt gggcggcatg accgggaccg gcaagaccga ggtgctggcc
361 cagctgggca acagcctgga tctcgaaggg ttggccaacc accgcggctc cagcttcggc
421 aagcgcgcca cgccgcagcc ctcgcagatc gatttcgaga acgcgctggc gatccgcctg
481 ctgaagatgc gcgcggcagg gatcgaaagc ttcgtgatcg aggacgaggc gcgcctggtc
541 gggcgctgct ccatcccgtt gccgctgttc cagggcatgc agcagtttcc gttgatctgg
601 ctggaggaca gtctgcaagg ccgggtcgag cgcatcctca aggactatgt gatcgacctg
661 tgcgccgagt tcatcagcga acagggtgag aacggtttcg ctgcattcgc tgagcgtctt
721 cagcaaagtc tggtcaatat caacaagcgg ctcggcggcg agcgttacca gcgcctggcc
781 ctgctgatgg atcaggccct ggcggagcag gcgcgcagcg gtgcggtcga tctgcatcgc
841 gcctggatcg aaacgttgtt gggcgaatat tacgacccga tgtacgccta ccaacgcgag
901 agcaaggccg agcgtatcga gttcgccggc gaccagccgt ctgtcgtcga gtatttgcgt
961 aaccgtaagg cgcgcagcgc gccgtaa
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptides encoded by Phage DNA & BLASTp search:

Frame 1: IPGRRSAPASSSMGRRRPSSLATSWSAAEPRTSGFRPGPISPGTIRRATSTVSVA
DFVRRSEFSCRDIPGARRQACGRTRVTS

[NO SIGNIFICANT SIMILARITY]

Frame 2: YRAEGRHLLQAAWADGGHRAWQPVGQRQNQGRADSGLGRFRPEQSAGLHLLFPWR
TSFADPNSPAGISRELVDKLAAALE

[NO SIGNIFICANT SIMILARITY]

Frame 3: TGQKVGTCFKQHGQTAAIELGNQLVSGRTKDERIQAWADFARNNPQGYIYCFRGG
LRSQIRILLQGYPGSSSTSLRPHSSN

LOCUS YP_001172420 328 aa linear BCT 16-NOV-2011
DEFINITION **tRNA 2-selenouridine synthase [Pseudom. Stutzeri A1501].**
ACCESSION YP_001172420 VERSION YP_001172420.1 GI:146282267

Score = 105 bits (262), Expect = 9e-32, Method: Compos.-based stats.
Identities = 46/58 (79%), Positives = 54/58 (93%), Gaps = 0/58 (0%)

```
Query  3   QKVGTCFKQHGQTAAIELGNQLVSGRTKDERIQAWADFARNNPQGYIYCFRGGLRSQI  60
           QKVGTC+KQHGQ AAIE+G+QLVSG+TK ER+QAWADFAR +P GY+YCFRGGLRS+I
Sbjct  7   QKVGTCYKQHGQKAAIEMGHQLVSGKTKAERLQAWADFARAHPDGYLYCFRGGLRSKI  64
```

***Pseudomonas stutzeri* gDNA library – Round 9 – Plaque F10**

ATACCGGGCAGAGGTCGGCACCTGCTTCAGCAGCATGGGCAGACGGCGGCCATCGAGCTTGGCAACCAGTTGG
TCAGCGGCAGAACCAAGGACGAGCGGATTCAGGCCTGGGCCGATTTCGCCCGGAACAATCCGCAGGGCTACAT
CTACTGTTTCCGTGGCGGACTTCGTTCGCAGATCCGAATTCTCCTGCAGGGATATCCCGGGAGCTCGTCGACA
AGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTT (up primer)

sequence is about 850bp long.
BLAST shows no significant similarity within *P. stutzeri* genome when BLAST search done without organism restriction and as high similarity search:

Peptide encoded by Phage DNA:

Frame 1: FRARTSISKD
Frame 2: SGLEPQLVKTKGRRKNRCSASIPL
Frame 3: QG

seemingly very short sequences considering the size of the insert?!

***Pseudomonas stutzeri* gDNA library – Round 9 – Plaque B11**

TTTTTCGTGCAGGAGGGCTGAGGTTCTTCGAGTCGGTGCGCAATGCCTATCTCGAGCGCGCGAAGCAGGCGCC
GCATCGCTACCGAGTGATCCGAATTCTCCTGCAGGGATATCCCGGGAGCTCGTCGACAAGCTTGCGGCCGCAC
TCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTA (up primer)

sequence abruptly ends at 196bp - too little primer or PCR product

2 main hits:

LOCUS NC_004777 37555 bp DNA linear PHG 26-MAR-2010
DEFINITION **Yersinia pestis phage phiA1122, complete genome.**
ACCESSION NC_004777 VERSION NC_004777.1 GI:30387453

Score = 73.1 bits (39), Expect = 3e-09 Identities = 44/46 (96%), Gaps = 1/46 (2%) Strand=Plus/Plus
Query 152 TAACTAG-TTAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTT 196
||||||| |||||||||||||||||||||||||||||||||||
Sbjct 21777 TAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTT 21822

and

LOCUS NC_001604 39937 bp DNA linear PHG 26-MAR-2010
DEFINITION **Enterobacteria phage T7, complete genome.**
ACCESSION NC_001604 VERSION NC_001604.1 GI:9627425

Score = 73.1 bits (39), Expect = 3e-09 Identities = 44/46 (96%), Gaps = 1/46 (2%) Strand=Plus/Plus

Query 152 TAACTAG-TTAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTT 196
||||||| |||||||||||||||||||||||||||||||||||
Sbjct 24160 TAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTT 24205

BLASTP search:

Frame 1: FFVQEG

[NO SIGNIFICANT SIMILARITY]

Frame 2: FSCRRAEVLRVGAQCLSRAREAGAASLPSDPNSPAGISRELVDKLAAALE

LOCUS YP_001173063 430 aa linear BCT 16-NOV-2011
DEFINITION **flagellar biosynthesis regulator FlhF [Pseudomonas stutzeri A1501].**
ACCESSION YP_001173063 VERSION YP_001173063.1 GI:146282910

Score = 21.2 bits (43), Expect = 0.077, Method: Comp. matrix adjust. Identities = 9/14 (64%), Positives = 12/14 (86%), Gaps = 0/14 (0%)

Query 34 PAGISRELVDKLAA 47
PA +SR L+DK+AA
Sbjct 167 PAELSRSLLDKVAA 180

Frame 3: FRAGGLRFFESVRNAYLERAKQAPHRYRVIRILLQGYPGSSSTSLRPHSSN

LOCUS YP_001173109 210 aa linear BCT 16-NOV-2011
DEFINITION **thymidylate kinase [Pseudomonas stutzeri A1501].**
ACCESSION YP_001173109 VERSION YP_001173109.1 GI:146282956

Score = 45.4 bits (106), Expect = 8e-11, Method: Compos.-based stats. Identities = 20/30 (67%), Positives = 25/30 (83%), Gaps = 0/30 (0%)

Query 1 FRAGGLRFFESVRNAYLERAKQAPHRYRVI 30
F G+RFFE+VR+AYL+RA+ AP RYRVI
Sbjct 155 FEQEGMRFFEAVRSAYLQRAEAAPSRYRVI 184

***Pseudomonas stutzeri* gDNA library – Round 15 – Plaque B7**

CGGAATCTTCGATCCGGTACAGGCAGGATTCGTTATCGGCAACTGGTGGAACTGGTGCTGCAGACACGGGAAC
GCCTGGCCACGCTGTATGCGTCCGAGCTGCCGCCAGCGGAGCTGCGCCGTCTCAAACAGGCAGAATTCATGCA
CCTGCGGCAGGGCTATCGAGCGCTGCGTGACGGCGCCTGGGAAGGTGATCCGAATTCTCCTGCAGGGATATCC
CGGGAGCTCGTCGACAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTCTT
GAGGGGTTAAGTTACTCGAGTGCGCCAGCTTGTCGACAGCTCCCGGGATATCCCTGCAGGAGAATTCGGATCA
CCTTCCCAGGCGCCGTCACGCAGCGCTGATAGCCCTGCCGCAGGTGCATGAATTCTGCCTGTTTGAGACGGCG
CAGCTCCGCTGGCGGCAGCTCGGACGCATACGCGTGGCCGGCGTTCGCGTGGCTGCAGCACCAGTTCCACCAG
TTGCCGATACGAATCCGCCCTGTTGATCCCCGAGCATCACACCTGACTGGAATACGACGCTCC (up primer)

TCGCAATCATGCTACTGTTCTCGAGTGCGGCCGCAGCTTGTCGACGAGCTCCCGGGATATCCCTGCAGGAGAA
TTCGGATCACCTTCCCAGGCGCCGTCACGCAGCGCTCGATAGCCCTGCCGCAGGTGCATGAATTCTGCCTGTT
TGAGACGGCGCAGCTCCGCTGGCGGCAGCTCGGACGCATACAGCGTGGCCAGGCGTTCCCGTGTCTGCAGCAC
CAGTTCCACCAGTTGCCGATAACGAATCCTGCCCTGTTGATCCCCGAGCATCACACCTGACTGGAATACGACA
GCTCCCTAGTTACTCGGGTGGGCCCGCTTGTCGACGAGCTCCCGGGATATGCTGCAGGAGAATTGGATCACCT
TCCCAGGCGCCGTCACGCAGCGCTCGATAGCCCTGCCGCAGGTGCATGAATTCTGCCTGTTTGAGACGGAGCA
GCTCCGCTGGCGGCAGCTCGGACGCATACAGCGTGGCCAGGCGTTCCCGTGTCTGCAGCACCAGTTCCACCAG
TTGCCGATACGAATCCTGCCCTGTTGATCCCCGAGCATCACACCTGACTGGAATACGACAGCTCCA (down primer)

LOCUS NC_015740 4547930 bp DNA circular BCT 13-FEB-2012
DEFINITION **Pseudomonas stutzeri ATCC 17588 = LMG 11199 chromosome, complete genome.**
ACCESSION NC_015740 VERSION NC_015740.1 GI:339492077

Features in this part of subject sequence: **aminopeptidase**
Score = 98.7 bits (108), Expect = 2e-19 Identities = 118/158 (75%), Gaps = 2/158 (1%) Strand=Plus/Minus

```
Query  31       CGTTATCGGCAACTGGTGGAACTGGTGCTGCAGACACGGGAACGCCTGGCCACGCTGTAT  90
                ||||||||||| |||||| ||||| ||||  | || || || || || ||| ||||||||
Sbjct  4441401  CGTTATCGGCAGCTGGTGCAACTGCTGCTCGACACCCGCGAGCGTCTTGCCTCGCTGTAT  4441342
Query  91       GCGTCCGAGCTGCCGCCAGCGGAGCTGCGCCGTCTCAAACAGGCAGAATTC-ATGCACCT  149
                || || | || |||| | ||||| |||    | ||||||||||| || |||  |||  ||
```

```
Sbjct  4441341  GCCTCGGGGCGGCCGGCGGCGGAACTGGAGGGCCTCAAACAGGCCGAGTTCGCTGC-GCT  4441283
Query  150      GCGGCAGGGCTATCGAGCGCTGCGTGACGGCGCCTGGG  187
                |||  ||| |||||||  | || ||||| ||  | ||||
Sbjct  4441282  GCGCCAGCGCTATCGCACCCTACGTGATGGTTCGTGGG  4441245
```

Sequence pos. 255 - 300 matches perfectly with various expression vectors.

***Pseudomonas stutzeri* gDNA library – Round 15 – Plaque B9**

GAGTATTCGAATCCGCGACGATTCAGCTATCAGACCTAAAATCAAGCCTACATATGTTTCCTATATATCAAAG
AAAAATATATTATTCACCACTACATAGTTGGGGAAAATATCACACATTTGCTATAACTAGAGGCTTACAGTGG
GTTACCTTATATCACTAAAAATACTATCTGAGTTTTATATAATGATTCTAAATACTAACCTACTGGCATAGTT
ACACACATTGACTCTATCTTGGCTGATGTTGCTAAAGATTTTAGGTCTTGATTCTGTCTCTGACAGATGTAAG
ATTCTGTGTCTGTTGTTTGCCATGGAGACTGAGGCTACATAATGTTAAGTTAGCCGTGTGTGTGTGTGTGTGT
GTGTGTGTATAGGAAGAGTTTTTTTTTCTTTTTCTTTTTTTTTTTTTTTTGAAACGGAGTCTCCCTCTGTTGC
CCAGGGTGG (up primer)

GCTTTGAAGTTCATGATACTGTTCTCGAGTGCGGCCGCAAGCTTTTAGGGCTTCAGAAGAGAGAAAGAAAGAG
GTGGAGGGTATTTTGCAGGTTTGCTTTCTTTGCTACTCTCCACTTTATGTTTCTTTCAATCTTCTCTAAATGT
GTCCCATTCCCTACCTCTACTCTCTGTGAAGAGGTGAAGAGAGAAGAGAAAGAAGACAGAAATGTGGACAGCA
TTCAGAGTGGATTCTTCACACCGTCATGCTTCCCCACTCACCCACAGGTGAAATGTGAAAGTACAGGATTCAT
TACTAGCAATATAGCCTCCAAACTCAGAACTGTAAATACACACAATTCCCTGTAGTCATCAGGTCTCAGCAAA
TTCATGTTAGCTGCTGAGTCAATGATTAATCAAAATTAATTAGGACCCTCATACTATAGAGAACAAAACTCCT
GCCTAAGAAATTTAGCTCAGCTAACTGTGAATTATCAAGCCAAGAATTTGGAAGGGGAGAGGGCAGGAAAGGT
TTTTGTAGGACAAGTAATATCACACATATTTAGAAAGGCTCCCAGCCAGGCACGGTGGCTCACGCCTGTAATC
CCAGCACTTTGGGAGGCCGAGGCAGGCAGATTTCCTGAGGTCAGGAGTTCAAGAGCAGCCTGGCCAATGTGGT
GAAACCCCATCTCTACTAAAAATGCAAAAACTAACTGGGCGTGGTGGCAGATGCCTGTAATCCCAGCTACTTG
GGAGGCTGAGGTAGGAGAATTGCTTGAACCCAGGAGACGAAGGTTGCAGTGAGCCAAGATCACGCCACTGCAC
TCCAGCCTGGGCAACAGAGCGAGACTCCGTTCTCAAAAAAAAAAAAAAAAGAAAAAGAAAAAAATACCTCTTC
(down primer)

BLAST search extended to "nucleotide collection", all organisms, discontiguous megablast

LOCUS NG_012566 1207825 bp DNA linear PRI 26-MAR-2012
DEFINITION **Homo sapiens interleukin 1 receptor accessory protein-like 2 (IL1RAPL2); and testis expressed 13A (TEX13A), RefSeqGene on chromosome X.**
ACCESSION NG_012566 VERSION NG_012566.2 GI:375151552

Query pos. 28 - 447 match DNA sequence pos. 1199580 - 1199144

Gene in backwards

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: SAIRPKIKPTYVSYISKKNILFTTT
Frame 2: QLSDLKSSLHMFPIYQRKIYYSPLHSWGKYHTFAITRGLQWVTLYH
Frame 3: SYQT

***Pseudomonas stutzeri* gDNA library – Round 15 – Plaque C7**

GGGCACTATATTCGATCCGGACGCTCGCCGCGCCGCAGCGCCCCCAGCAGCTGGTCCTTGGCCTTGGGGACGA
TCACCTGCCTGGTGCTGGGCATGGGCATCCCGACCATCCCCAACTACATCATCACCAGCGCCATCGCCGCACC
GGCACTGGAGCAGCTGGGTGTGCCGCTGATCAGTTTTGTTTTTACCGGGATGCTGATGACTACGGTGGCGGTG
GCGGGAACCGAGCATGACGAGCGACGTGTCGATCCGAATTCTCCTGCAGGGATATCCCGGGAGCTCGTCGACA
AGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTCTTGGAGGGGTTA (up primer)

TACACGTATCATGCTACTGTTCTCGAGTGCGGCCGCAGCTTGTCGACGAGCTCCCGGGATATCCCTGCAGGAG
AATTCGGATCGACACGTCGCTCGTCATGCTCGGTTCCCGCCACCGCCACCGTAGTCATCAGCATCCCGGTAAA
AACAAAACTGATCAGCGGCACACCCAGCTGCTCCAGTGCCGGTGCGGCGATGGCGCTGGTGATGATGTAGTTG
GGGATGGTCGGGATGCCCATGCCCAGCACCAGGCAGGTGATCGTCCCCAAGGCCAAGGACCAGCTGCTGGGGG
CGCTGCGGCGCGGCGAGGGCGATCCCCGAGCATCACACCTGACTGGAATACGACAGCTCCT (down
primer)

LOCUS CP002622 4689946 bp DNA circular BCT 17-JUN-2011
DEFINITION **Pseudomonas stutzeri DSM 4166, complete genome.**
ACCESSION CP002622 VERSION CP002622.1 GI:327478612

Features in this part of subject sequence: **transporter, putative**
Score = 134 bits (148), Expect = 2e-30 Identities = 91/102 (89%), Gaps
= 0/102 (0%) Strand=Plus/Plus

in frame with protein of "transporter, putative"

```
Query  76       ACCTGCCTGGTGCTGGGCATGGGCATCCCGACCATCCCCAACTACATCATCACCAGCGCC  135
                |||||||||||||||||||||||||||||||||| |||||||||||||||||||| |
Sbjct  4479683  ACCTGCCTGGTGCTGGGCATGGGCATCCCGACCATTCCCAACTACATCATCACCAGCTCG  4479742
Query  136      ATCGCCGCACCGGCACTGGAGCAGCTGGGTGTGCCGCTGATC  177
           |||||||||||| || |||  | | || || ||||||||||||
Sbjct  4479743  ATCGCCGCACCCGCGCTGCTGGATCTCGGCGTGCCGCTGATC  4479784
```

***Pseudomonas stutzeri* gDNA library – Round 15 – Plaque H7**

AACCGGATTTCGGCCAAGACCGCATCAGGCTCACGCCGGGCGGCGGATTGTCACGCAGCTGGTCGCCCAGCTG
ACGGCCGATGCCATAGCTTACGCGCGTTTCGTCGGTGGACAGATTAAGGTCGGTCATGCCTGGGCTCCGCTCT
GGAAAAAGGTCGCACAGCTTAGGGGTGCGACCGTTTCGTCGCAAGCCGCGGACGAAACCGCAACAGCCCCTGG
CACAGCGTACCCGGCGAACCGCCTCTCAATGTTTGGTGAGTTTGTCCAGATAGCCCATGGCAAAGGCCGAGAA
AACGAAGGTCAGGTGAATGATCCGAATTCTCCTGCAGGGATATCCCGGGAGCTCGTCGACAAGCTTGCGGCCG
CACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTTTGAGGGGTTAAGTTACTCGAGTGCGCGCAA
GCTTGTCGACGAGCTCCCGGGATATCCCTGCAGGAGAATTCGGATCATTCACCTGACCTTCGTTTTCTCGGCC
TTTGCCATGGGCTATCTGGACAAACTCACCAAACATTGAGAGGCGGTTCGCCGGGTACGCTGTGCCAGGGGCT
GTTGCGGTTTCGTCCGCGGCTTGCGACGAAACGGTCGCGCCCCTAAGCTGTGCGACCTTTTTCAGAGCGGAGC
CCAGCATGACGACCTTAATCTGTCACCGACGAAACGCGCGTAAGCTATGGCATCGGCCGTCAGCTGGGCGACC
AGCTGCGTGACAATCCGCCGCCAGCGTGAGCCTGGATGCGGTGATCCGGGAGCTCACACCTGACTGGAATACC
ACCGCTCCCA (up primer)

GTGACGTTTTCATAACTCTGCATCTCGAGTGCGGCCGCAAGCTTGTCGACGAGCTCCCGGGATATCCCTGCAG
GAGAATTCGGATCATTCACCTGACCTTCGTTTTCTCGGCCTTTGCCATGGGCTATCTGGACAAACTCACCAAA
CATTGAGAGGCGGTTCGCCGGGTACGCTGTGCCAGGGGCTGTTGCGGTTTCGTCCGCGGCTTGCGACGAAACG
GTCGCACCCCTAAGCTGTGCGACCTTTTTCCAGAGCGGAGCCCAGGCATGACCGACCTTAATCTGTCCACCGA
CGAAACGCGCGTAAGCTATGGCATCGGCCGTCAGCTGGGCGACCAGCTGCGTGACAATCCGCCGCCCGGCGTG
AGCCTGGATGCGGTGATCCCCGAGCATCACACCTGACTGGAATACGACAGCTCCAAGTTACTCGATTGGGCGC
AGCTTGTGACGAGCTCCCGGGATATCCCTGCAGGAGAATTCGGATCATTCACCTGACCTTCGTTTTCTCGGCC
TTTGCCATGGGCTATCTGGACAAACTCACCAAACATTGAGAGGCGGTTCGCCGGGTACGCTGTGCCAGGGGCT
GTTGCGGTTTCGTCCGCGGCTTGCGACGAAACGGTCGCACCCCTAAGCTGTGCGACCTTTTTCAGAGCGGAGC
CCAGGCATGACCGACCTTAATCTGTCCACCCACGAAACGCGCGTAAGCTATGGCATCGGCCGTCAGCTGGGGC
GACCAGCTGCGTGACAATCCGCCGCCCGGCGTGAGCCTGGATGCGGTGATCCGCGAGCCTTCCACCTGACTGG
AATAACACAGCTC (down primer)

very similar (alignment) with D-PS-H7

DEFINITION **Pseudomonas stutzeri A1501 chromosome, complete genome.**
ACCESSION NC_009434 VERSION NC_009434.1 GI:146280397

Features in this part of subject sequence:
peptidyl-prolyl cis-trans isomerase, FKBP-type
Score = 165 bits (182), Expect = 3e-39 Identities = 129/150 (86%),
Gaps = 3/150 (2%) Strand=Plus/Minus

```
Query  11       CGGCCAAGACCGCATC-AGGCTCACGCCGGGCGGCGGATTGTCACGCAGCTGGTCGCCCA  69
                ||||  |  ||||| || ||||| || || |||||||| |||||||||||||| || ||||
Sbjct  1059240  CGGCGATCACCGCGTCGAGGCTGACACCCGGCGGCGGGTTGTCACGCAGCTGATCACCCA  1059181

Query  70       GCTGACGGCCGATGCCATAGCTTACGCGCGTTTCGTCGGTGGACAGATTAAGGTCGGTCA  129
                |||||||||||||||| ||||| ||||| |||||||||||||| ||||| || |||||||
Sbjct  1059180  GCTGACGGCCGATGCCGTAGCTGACGCGGGTTTCGTCGGTGGAGAGATTGAGTTCGGTCA  1059121

Query  130      TGCCTGGGCTCCGCTCTGG-AAAAAGGTCG  158
                |||| |||||||| || || ||||||||||
Sbjct  1059120  TGCC-GGGCTCCGGTCAGGTAAAAAGGTCG  1059092
```

Features in this part of subject sequence:
peptidyl-prolyl cis-trans isomerase, FKBP-type
Score = 149 bits (164), Expect = 2e-34 Identities = 123/145 (85%), Gaps = 5/145 (3%) Strand=Plus/Plus

```
Query  637      CGACCTTTTT--CAGAGCGGAGCCCAGCATGAC-GACCTTAATCTGTC-ACCGACGAAAC  692
                |||||||||  | || |||||||| ||||||| || || ||||| || ||||||||||
Sbjct  1059092  CGACCTTTTTACCTGACCGGAGCCCGGCATGACCGAACTCAATCTCTCCACCGACGAAAC  1059151

Query  693      GCGCGTAAGCTATGGCATCGGCCGTCAGCTGGGCGACCAGCTGCGTGACAATCCGCCGCC  752
                ||||| ||||| |||||||||||||||||| || |||||||||||||| ||||||||
Sbjct  1059152  CCGCGTCAGCTACGGCATCGGCCGTCAGCTGGGTGATCAGCTGCGTGACAACCCGCCGCC  1059211

Query  753      -AGCGTGAGCCTGGATGCGGTGATC  776
                | || ||||| || |||||||||
Sbjct  1059212  GGGTGTCAGCCTCGACGCGGTGATC  1059236
```

Features in this part of subject sequence:
hypothetical protein Score = 73.4 bits (80), Expect = 1e-11
Identities = 58/70 (83%), Gaps = 0/70 (0%) Strand=Plus/Plus

```
Query  481      GATCATTCACCTGACCTTCGTTTTCTCGGCCTTTGCCATGGGCTATCTGGACAAACTCAC  540
                |||||| ||||| ||||||||  | ||||||||||||||||| ||  |||||||  ||||
Sbjct  1056890  GATCATCCACCTCACCTTCGTGGTGTCGGCCTTTGCCATGGGTTACATGGACAAGATCAC  1056949

Query  541      CAAACATTGA  550
                ||| || |||
Sbjct  1056950  CAAGCACTGA  1056959
```

Features in this part of subject sequence:
hypothetical protein Score = 73.4 bits (80), Expect = 1e-11
Identities = 58/70 (83%), Gaps = 0/70 (0%) Strand=Plus/Minus

```
Query  245      TCAATGTTTGGTGAGTTTGTCCAGATAGCCCATGGCAAAGGCCGAGAAAACGAAGGTCAG  304
                ||| || |||||||  |||||||  || |||||||||||||||  |  |||||||| ||
Sbjct  1056959  TCAGTGCTTGGTGATCTTGTCCATGTAACCCATGGCAAAGGCCGACACCACGAAGGTGAG  1056900

Query  305      GTGAATGATC  314
                ||| ||||||
Sbjct  1056899  GTGGATGATC  1056890
```

8.3.3.4 Artesunate

8.3.3.4.1 Human colon

Library did not converge.

8.3.3.4.2 Human Alzheimer's brain

Alzheimer's Brain cDNA Library – Round 12 – Plaque B9

TGGGGGGGTTCGGGGCAC**GAATTCA**AGCCGAGATCGGGCTTGGGCCCAGAGCATGTTCCAGATCCCAGAGTTT
GAGCCGAGTGAGCAGGAAGACTCCAGCTCTGCAGAGAGGGGCCTGGGCCCCAGCCCCGCAGGGGACGGGCCCT
CAGGCTCCGGCAAGCATCATCGCCAGGCCCCAGGCCTCCTGTGGGACGCCAGTCACCAGCAGGAGCAGCCAAC
CAGCAGCAGCCATCATGGAGGCGCTGGGGCTGTGGAGATCCGGAGTCGCCACAGCTCCTACCCCGCGGGGACG
GAGGACGACGAAGGGATGGGGGAGGAGCCCAGCCCCTTTCGGGGCCGCTCGCGCTCGGCGCCCCCCAACCTCT
GGGCAGCACAGCGCTATGGCCGCGAGCTCCGGAGGATGAGTGACGAGTTTGTGGACTCCTTTAAGAAGGGACT
TCCTCGCCCGAAGAGCGCGGGCATAGCAACGCAAATGCGGCAAAGCTCCAGCTGGACGCGAGTCTTCCAGTCC
TGGTGGGATCGGAACTTGGGCAGGGGAAGCTCCGCCCCCTCCCAGTGACCTTCGCTCCACATCCCGAAACTCC
ACCCGTTCCCACTGCCCTGGGCAGCCATCTTGAATATGGGCGGAAGTACTTCCCTCAGGCCTATGCAAAAAGA
GGATCCGTGCTGTCTCCTTTGGAGGGAGGGCTGACCCAGATTCCCTTCCGGTGCGTGTGAAGCCACGGAAGGC
TTGGTCCCATCGGAAGTTTTGGGTTTTCCGCCCACAGCCGCCGGAAGTGGCTCCGTGGCCCCGCCCTCAGGCT
CCGGGCTTTCCCCCAGGCGCCTGCGCTAAGTCGCGAGCCAGGTTTAACCGTTGCGTCACCGGGACCCGAGCCC
CGCGATGCCCTGGGGGCCGTGCTCACTACCAAATGTTAATAAAGCCCGCGTCTGTGCCAAAAAAAAGCTTGCG
GCCGCACTCGAGTACTAGTTAACCCCTTGGGGCCTCTAAAC (up primer)

TTTTTTTTGGCACAGACgCGGGCTTTATTAACATTTGGTAGTGAGCACGGCCCCCAGGGCATCGCGGGGGCTC
GGGTCgCGGTGACCCAACGGTTAAACCTGGCTCGCGACTTAGCGCAGGCGCCTGGGGGAAAGCCCGGAGCCTG
AGGGCGGGGCCACGGAGCCACTTCCGGCGGCTGTGGGCGGAAAACCCaAAACTTCCGATGGGACCAAGCCTTC
CGTGGCTTCACACGCACCGGAAGGGAATCTGGGTCAGCCCTCCCTCCaaaGGAGACAgCACGGATCCTCTTTT
TGCATAGGCCTGAGGGAAGTACTTCCGCCCATATTCAAGATGGCTGCCCAGGGCAGTGGGAACGGGTGGAGTT
TCGGGATGTGGAGCGAAGGTCACTGGGAGGGGGCGGAGCTTCCCCTGCCCAAGTT
CCGATCCCACCAGGACTGGAAGACTCGCGTCCAGCTGGAGCTTTGCCCCATC (down primer)

LOCUS NM_032989 986 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens BCL2-associated agonist of cell death (BAD), transcript variant 2, mRNA.**
ACCESSION NM_032989 VERSION NM_032989.2 GI:197116382

```
  1 aactagggcc cggagcccgg ggtgctggag ggaggcggca ggcccgggtc aggggcctcg
 61 agatcgggct tgggcccaga gcatgttcca gatcccagag tttgagccga gtgagcagga
121 agactccagc tctgcagaga ggggcctggg ccccagcccc gcaggggacg ggccctcagg
181 ctccggcaag catcatcgcc aggccccagg cctcctgtgg gacgccagtc accagcagga
241 gcagccaacc agcagcagcc atcatggagg cgctggggct gtggagatcc ggagtcgcca
301 cagctcctac cccgcgggga cggaggacga cgaagggatg ggggaggagc ccagcccctt
361 tcggggccgc tcgcgctcgg cgccccccaa cctctgggca gcacagcgct atggccgcga
421 gctccggagg atgagtgacg agtttgtgga ctcctttaag aagggacttc ctcgcccgaa
481 gagcgcgggc acagcaacgc agatgcggca aagctccagc tggacgcgag tcttccagtc
541 ctggtgggat cggaacttgg gcaggggaag ctccgccccc tcccagtgac cttcgctcca
601 catcccgaaa ctccacccgt tcccactgcc ctgggcagcc atcttgaata tgggcggaag
661 tacttccctc aggcctatgc aaaaagagga tccgtgctgt ctcctttgga gggagggctg
721 acccagattc ccttccggtg cgtgtgaagc cacggaaggc ttggtcccat cggaagtttt
781 gggttttccg cccacagccg ccggaagtgg ctccgtggcc ccgccctcag gctccgggct
841 ttcccccagg cgcctgcgct aagtcgcgag ccaggtttaa ccgttgcgtc accgggaccc
901 gagccccccgc gatgccctgg gggccgtgct cactaccaaa tgttaataaa gcccgcgtct
961 gtgccgccga aaaaaaaaaa aaaaaa //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = complete translation of CDS):

Frame 1: **SRDRAWAQSMFQIPEFEPSEQEDSSSAERGLGPSPAGDGPSGSGKHHRQAPGL LWDASHQQEQPTSSSHHGGAGAVEIRSRHSSYPAGTEDDEGMGEEPSPFRGRS RSAPPNLWAAQRYGRELRRMSDEFVDSFKKGLPRPKSAGIATQMRQSSSWTRV FQSWWDRNLGRGSSAPSQ**
Frame 2: AEIGLGPRACSRSQSLSRVSRKTPALQRGAWAPAPQGTGPQAPASIIARPQAS CGTPVTSRSSQPAAAIMEALGLWRSGVATAPTPRGRRTTKGWGRSPAPFGAAR ARRPPTSGQHSAMAASSGG
Frame 3: PRSGLGPEHVPDPRV

Alzheimer's Brain cDNA Library – Round 12 – Plaque A10

TCTTTTAGGGCTAGGAC**GAATTCA**TGTTTCAGGAGCCTAAGACCCTCCCAGAGCCCAGGGGCTTCACCGCAGA
CCCGAAGCCATTGAGCACATCACCCAAAGCAGTGGCCAACATCGCGGACCCCTGTGCCTTGTCACAGATGGGT
GCTGGTCCTCAGGCGTTGGGGACACTGCTGGGTCGATGGGGTCGGATTCTGCCAGTTTCTGCTCTGCAGCCAA
AGATGGTCAGAAGCATTGTCACTTCAGTAACATCAAGTGCTCAAAGACATGGCAACCGTTCAGTGGTACTTAA
GTATTCAAAATATACAACTACAGATTCTCTGACAGAAACCAGCACGGGGTCTTCACCTTCATTCACCCCACAG
GCGACATGCGAGGGAGAACAGCATCTCAGTGGTGATTTCCAAACCAAGCCTTTGTTTTCGGTGTGGGGTTTTG
GGGGTTTGCTTTAATGTTTTTGAAATTGTAAATGTTGGGCTTTGTATTTTGATGTAAACTGAGCATAATGGCA
TTTTAGGGCCTGTGACCAAAAATGAAGCTTGTAACGACCATGGATCTGAATAAACATGTCCTTGCTTCTGAGT
CTTCTGAAAAAAAAAAAAAAAAAAAAAAAAAAACTTGGGGCCCCCCTCCAAAAAAAAATTAACCCCTTGGGGC
CCTCAAAAGGGGTCTGGGGGGGGTTTTAA (up primer)

LOCUS NM_017951 3920 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens sphingomyelin phosphodiesterase 4, neutral membrane (neutral sphingomyelinase-3) (SMPD4), transcript variant 2, mRNA.**
ACCESSION NM_017951 VERSION NM_017951.4 GI:300192967

```
   1 ctaaggacca ctccctcccc cgcactcctg cctcgccatt tctcttcccc gcccggccgg
  61 ccttcgcttt gcgcacgcgc cttttgaggt aacggcccaa agaggtggaa gcgcttttcc
 121 cgcccggccg cggggcgtgg ctctgcgcgc agcttgatga cgactttcgg cgccgtggcg
 181 gaatggcggc ttccatctct gaggcgagcg acgctatgga tcccacagtg gtttgctaag
  //
2701 ctctacgcct ctgccatgac actgctgacc gagcgggggа agctgcacca gccctgaagg
2761 tgtcagctgc cttcagagca ggctggaggg atttgccaca cagccccacc cttgggctga
  //
3301 tcttcatgtt ctcccagctg ttccaagact gggccgtaga attccatgtt tcaggagcct
3361 aagaccctcc cagagcccag gggcttcacc gcagacccca agccattgag cacatcaccc
  //
3841 gcctgtgacc aaaaatgaag cttgtaacga ccatggatct gaataaacat gtccttgctt
3901 ctgagtcttc tg gcacctgg //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame):

Frame 1: SCFRSLRPSQSPGASPQTRSH
Frame 2: HVSGA
Frame 3: MFQEPKTLPEPRGFTADPKPLSTSPKAVANIADPCALSQMGAGPQALGTLL GRWGRILPVSALQPKMVRSIVTSVTSSAQRHGNRSVVLKYSKYTTTDSLT ETSTGSSPSFTPQATCEGEQHLSGDFQTKPLFSVWGFGGLL

BLASTp search:

Frame 1 protein:
- perilipin
  query cov 61%, E = 8e-04, max ident 77%
- **mediator complex subunit MED25 variant MED25_i4**
  query cov 71%, E = 2e-03, max ident 100%

Frame 3 protein:
- *no significant similarities found (discontiguous megablast)*

Comment: there are multiple hits all matching sphingomyelin phosphodiesterase 4, neutral membrane, but different variants, some described as non-coding RNA, others with CDS.

Alzheimer's Brain cDNA Library – Round 12 – Plaque C10

CGCAAATACGGTCGCGAC**GAATTCA**TCTTGGGGTCCCCGAAGCAACCTGGAATCAGCAAGATGGAAGTCCTGG
GGGTTCGTGGCCAAACTGCATAAGAAGGAACCGCAGCACTGGCCAGTGCAGTACCGTGAGGCCCTAGCAGACG
AGGCCGACAGGGCCAGAGCCAAGGCCAGAGCTGAAGCCAGTATGAGGGCCAGGGCCAGTGCTAGGGCCGGCAT
CCACCTCTGGTGAGGGTTGGTGAAAAGTTGGCCAGTGGGTCCCCGTGAGGACGAACTACTGTCCTGAGTCATA
AGTAATATGGGTGGGGCGAGGGTCTTATTTCTGTAGAAATCGTGTGACTTTAAGGATTTAGATTTTGTATCTT
ATGTTTTGTAACATTTAATAATTACTGTTAAAATGCTGTTTGTAAATGAGATTGGTCTACTTTTTCCTGTAGG
ATTTTATTGTAGAGTTTTGCTGGTTTTGTAAAATGGATGGAAGAACTTTGTATTTATACTGTGATTTTGAACA
GATTATGCAACATTGGAAGGAAGGCTGTACTTTGATGGTTTGAAGGAACTCAGCAGTATGATGATCTGGTTCC
AGGGGAAAAAAATAGCTGGTTGGTGTCTAGCCCCCCAACACTTTTGTCTGTTGTGTATAAAAGAAGAAAGACT
GGCATGTACCTTCATTTGCTTAGCTATTTGAGTATCTAGAGAAAAATTAAAATGCAATGAGTTAGCAGTATAC
CCTGGCACACTTAATAAATTAAACATTTGTGGAGCAAAAAAAAAGCTTGCGGCCGCACTCGAGTAACTAGTTA
ACCCCTTGGGGCCTCTAAACGGTCTTTGGAGGGGTTAA (up primer)

CGGGTTACTGTTCTCGAGTGCGGCCGCAAGCTTTCGGGATGTGGAGCGAAGGTCACTGGGAGGGGGCGGAGCT
TCCCCTGCCCAAGTTCCGATCCCACCAGGACTGGAAGACTCGCGTCCAGCTGGAGCTTTGCCGCATCTGCGTT
GCTGTGCCCGCGCTCTTCGGGCGAGGAAGTCCCTTCTTAAAGGAGTCCACAAACTCGTCACTCATCCTCCGGA
GCTCGCGGCCATAGCGCTGTGCTGCCCAGAGGTTGGGGGGCGCCGAGCGCGAGCGGCCCCGAAAGGGGCTGGG
CTCCTCCCCCATCCCTTCGTCGTCCTCCGTCCCCGCGGGGTAGGAGCTGTGGCGACTCCGGATCTCCACAGCC
CCAGCGCCTCCATGATGGCTGCTGCTGGTTGGCTGCTCCTGCTGGTGACTGGCGTCCCACAGGAGGCCTGGGG
CCTGGCGATGATGCTTGCCGGAGCCTGAGGGCCCGTCCCCTGCGGGGCTGGGGCCCAGGCCCCTCTCTGCAGA
GCTGGAGTCTTCCTGCTCACTCGGCTCAAACTCTGGGATCTGGAACATGCTCTGGGCCCAAGCCCGATCTCGA
GGCCCCTGACCCGGGCCTGCCGCCTCGCTTGAATTCGGATCCCCGAGCATCACACC
TGACTGGAATACGACAGCTCCAGAGG (down primer)

LOCUS NM_0221491700 bp mRNA linear PRI 15-AUG-2011
DEFINITION **Homo sapiens melanoma antigen family F, 1 (MAGEF1), mRNA.**
ACCESSION NM_022149 VERSION NM_022149.4 GI:58530878

```
   1 gcgggcgcgg actgaggctg cgcgccgcag gttccggctg ctggcggcgt tgcggccgca
  //
 181 acacccctcg gcccgtgtac gcgctctgca cctgcctgcc cgaaaac atg ttgcagacac
 241 cagagagcag ggggctcccg gtcccgcagg ccgaggggga gaaggatggc ggccatgatg
  //
 901 gttacaggcg ggtgcctcac accaatccac cagaatatga attctcttgg ggtccccgaa
 961 gcaacctgga aatcagcaag atggaagtcc tggggttcgt ggccaaactg cataagaagg
1021 aaccgcagca ctggccagtg cagtaccgtg aggccctagc agacgaggcc gacagggcca
1081 gagccaaggc cagagctgaa gccagtatga gggccagggc cagtgctagg gccggcatcc
1141 acctctggtg agggttggtg aaaagttggc cagtgggtcc ccgtgaggac gaactactgt
  //
1621 aattaaaatg caatgagtta gcagtatacc ctggcacact taataaatta aacatttgtg
1681 gaaaaaaaaa aaaaaaaaaa //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame):

Frame 1: SSWGPRSNLEISKMEVLGVRGQTA
Frame 2: HLGVPEATWKSARWKSWGFVAKLHKKEPQHWPVQYREALADEADRARAKARAEASM
RARASARAGIHLW
Frame 3: ILGSPKQPGNQQDGSPGGSWPNCIRRNRSTGQCSTVRP

Alzheimer's Brain cDNA Library – Round 12 – Plaque D10

TAAGGCATCGGTCTAGAC**GAATTCA**TGTTTCAGGAGCCTAAGACCCTCCCAGAGCCCAGGGGCTTCACCGCAG
ACCCCAAGCCATTGAGCACATCACCCAAAGCAGTGGCCAACATCGCGGACCCCTGTGCCTTGTCACAGATGGG
TGCTGGTCCTCAGGCGTTGGGGACACTGCTGGGTCGATGGGGTCGGATTCTGCCAGTTTCTGCTCTGCAGCCA
AAGATGGTCAGAAGCATTGTCACTTCAGTAACATCAAGTGCTCAAAGACATGGCAACCGTTCAGTGGTACTTA
AGTATTCAAAATATACAACTACAGATTCTCTGACAGAAACCAGCACGGGGTCTTCACCTTCATTCACCCCACA
GGCGACATGCGAGGGAGAACAGCATCTCAGTGGTGATTTCCAAACCAAGCCTTTGTTTGTTTTCGGTGTGGGG
TTTTGGGGGTTTGCTTTAATGTTTTTGAAATTGTAAATGTTGGGCTTTGTATTTTGATGTAAACTGAGCATAA
TGGCATTTTAGGGCCTGTGACCAAAAATGAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCC
TCTAAACGGGTCTTGGAGGGGGT (up primer)

Aligns perfectly with A10, except:

position 60: G (A10) instead of C (D10)
position 406: GTTT (in A10) is missing in D10
position 526 and beyond: mostly mismatches

Alzheimer's Brain cDNA Library – Round 12 – Plaque E10

CTGCCATTGGGCCGAGAC**GAATTCA**TGTTTAGGAGCCTAAGACCCTCCCAGAGCCCAGGGGCTTCACCGCAGA
CCCCAAGCCATTGAGCACATCACCCAAAGCAGTGGCCAACATCGCGGACCCCTGTGCCTTGTCACAGATGGGT
GCTGGTCCTCAGGCGTTGGGGACACTGCTGGGTCGATGGGGTCGGATTCTGCCAGTTTCTGCTCTGCAGCCAA
AGATGGTCAGAAGCATTGTCACTTCAGTAACATCAAGTGCTCAAAGACATGGCAACCGTTCAGTGGTACTTAA
GTATTCAAAATATACAACTACAGATTCTCTGACAGAAACCAGCACGGGGTCTTCACCTTCATTCACCCCACAG
GTGACATGCGAGGGAGAACAGCATCTCAGTGGTGATTTCCAAACCAAGCCTTTGTTTTCGGTGTGGGGTTTTG
GGGGTTTGCTTTAATGTTTTTGAAATTGTAAATGTTGGGCTTTGTATTTTGATGTAAACTGAGCATAATGGCA
TTTTAGGGCCTGTGACCAAAAATGAAGCTTGTAACGACCATGGTTCTGAATAAACATGTCCTTGCTTCTGAAA
AAAAAAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTTTGGAAGGGGGTA
(up primer)

Aligns perfectly with A10, except:

position 60: G (in A10) instead of C (in E10)
position 350: C (A10) instead of T (E10)
position 538: A (A10) instead of T (E10)
position 565 and beyond: multiple mismatch

Alzheimer's Brain cDNA Library – Round 12 – Plaque D11

GAATTCAGCGGCAGATCTGGGCTCAGGGGAAGGAGGTGGAATAGTCCCAAAGCCAGCCGATGTTTCCAAGGTG
GCCACTGGGGACGGAGCCAGAATGGCCGCCCCTCCCCCAGAGCTCCCTGGGAAGTGGAACTTGAAGGATATTC
TGGGAAGGCGAATGGGGGAAGGGAGCCATCCATAAAAACAAAGTGGATATTCAGCCTAGGTCAACCCCGCCCC
TCTGGGTAGCAGCTATTTGGGTGACAGAACGGCCTAGGAAGAAACTGAAGGCCGGGCGCGGTGGCTCACGCCT
GTAATCCCAACACTTTGGGAGGCCGAGGCAGGTGGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAA
CAAGGAGAAACCCCCATCTCTACTAAAAAATGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTTACCGGGG
ATGGGGGGCACTGCCTTGTATCCCCACTTCCCCGGAAACTTAAGGGGGAAAAATCTTTGAACCTGGGAAGGCG
AAATTTCCAGGAACTTAAATTGGCCCCTTTGCCCCCCCCCCGGGGAAAAAAAAGAAAACCCCCTTCTAAAAAA
ACTTGCCGGCCCCCTTCAATAAATTATTAAACCCCTGGGGGCCTCTAAACGGTTTGGAGGGGGTTAAAACAAG
TCCCCGGGGGGGCCAGGTTTTTTAGAAAGGGATTTCCCTTTGGTGGCCAGGGCGGGAGGGCATGGGTCATTTC
ACCTAGGGAACTTGCATCCCAGGTTAAGCATTTTCCCCCCCGGGTCCAGATGCTGGAAATCAGGGTTTGCCCC
TGCGGGCAATTTCTTTGCTTTTTTTTTTTTTTTTCATATATAAAAGAGGGGGGTTTCTGTGTGAGGAGGGGTA
AAATCTCAGGAAAACACTCTGCCCAAGAGAGAAGAAGGAGGCGCCGCGCGCCATTTTCTCGAGGGTTGCGCCA
TGTGTCCCAGAGGGGGGTATATGTAACATCACTTTATATGATGCTGTCATCTCTCTCACACTACTATACTCTC

TCGGCGTAGAGAGGGCACTGTTCTCCGGCATGGAACACGATGGTGAATACTCATCGAATGAAGATGACGCATA
GTGAAGTTGAACTAG (up primer)

LOCUS AC087388 121017 bp DNA linear PRI 08-OCT-2002
DEFINITION **Homo sapiens chromosome 17, clone RP11-199F11, complete sequence.**
ACCESSION AC087388 VERSION AC087388.9 GI:23592178

8.3.3.4.3 Human colon tumour

Colon Tumour cDNA Library – Round 9 – Plaque B1

GTCTCGAGCAC**GAATTCA**AGCGATCGGGCTTGGGCCCAGAGCATGTTCCAGATCCCAGAGTTTGAGCCGAGTG
AGCAGGAAGACTCCAGCTCTGCAGAGAGGGGCCTGGGCCCCAGCCCCGCAGGGGACGGGCCCTCAGGCTCCGG
CAAGCATCATCGCCAGGCCCCAGGCCTCCTGTGGGACGCCAGTCACCAGCAGGAGCAGCCAACCAGCAGCAGC
CATCATGGAGGCGCTGGGGCTGTGGAGATCCGGAGTCGCCACAGCTCCTACCCCGCGGGGACGGAGGACGACG
AAGGGATGGGGGAGGAGCCCAGCCCCTTTCGGGGCCGCTCGCGCTCGGCGCCCCCCAACCTCTGGGCAGCACA
GCGCTATGGCCGCGAGCTCCGGAGGATGAGTGACGAGTTTGTGGACTCCTTTAAGAAAGGACTTCCTCGCCCG
AAGAGCGCGGGCACAGCAACGCAGATGCGGCAAAGCTCCAGCTGGACGCGAGTCTTCCAGTCCTGGTGGGATC
GGAACTTGGGCAGGGGAAGCTCCGCCCCCTCCCAGTGACCTTCGCTCCACATCCCGAAAGCTTGCGGCCGCAC
TCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTTTGGAGGGGTTA (up primer)

LOCUS NM_032989 986 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens BCL2-associated agonist of cell death (BAD), transcript variant 2, mRNA.**
ACCESSION NM_032989 VERSION NM_032989.2 GI:197116382

```
  1 aactagggcc cggagcccgg ggtgctggag ggaggcggca ggcccgggtc aggggcctcg
 61 agatcgggct tgggcccaga gcatgttcca gatcccagag tttgagccga gtgagcagga
121 agactccagc tctgcagaga ggggcctggg ccccagcccc gcaggggacg ggccctcagg
181 ctccggcaag catcatcgcc aggccccagg cctcctgtgg gacgccagtc accagcagga
241 gcagccaacc agcagcagcc atcatggagg cgctggggct gtggagatcc ggagtcgcca
301 cagctcctac cccgcgggga cggaggacga cgaagggatg ggggaggagc ccagcccctt
361 tcggggccgc tcgcgctcgg cgccccccaa cctctgggca gcacagcgct atggccgcga
421 gctccggagg atgagtgacg agtttgtgga ctcctttaag aagggacttc ctcgcccgaa
481 gagcgcgggc acagcaacgc agatgcggca aagctccagc tggacgcgag tcttccagtc
541 ctggtgggat cggaacttgg gcaggggaag ctccgccccc tcccagtgac ttcgctccca
601 catcccgaaa ctccacccgt tcccactgcc ctgggcagcc atcttgaata tgggcggaag
661 tacttccctc aggcctatgc aaaaagagga tccgtgctgt ctcctttgga gggagggctg
721 acccagattc ccttccggtg cgtgtgaagc cacggaaggc ttggtcccat cggaagtttt
781 gggttttccg cccacagccg ccggaagtgg ctccgtggcc ccgccctcag gctccgggct
841 ttcccccagg cgcctgcgct aagtcgcgag ccaggtttaa ccgttgcgtc accgggaccc
901 gagcccccgc gatgccctgg gggccgtgct cactaccaaa tgttaataaa gcccgcgtct
961 gtgccgccga aaaaaaaaaa aaaaaa //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = complete translation of CDS):

Frame 1: **SDRAWAQSMFQIPEFEPSEQEDSSSAERGLGPSPAGDGPSGSGKHHRQAPGLL WDASHQQEQPTSSSHHGGAGAVEIRSRHSSYPAGTEDDEGMGEEPSPFRGRSR SAPPNLWAAQRYGRELRRMSDEFVDSFKKGLPRPKSAGTATQMRQSSSWTRVF QSWWDRNLGRGSSAPSQ**
Frame 2: AIGLGPRACSRSQSLSRVSRKTPALQRGAWAPAPQGTGPQAPASIIARPQASC GTPVTSRSSQPAAAIMEALGLWRSGVATAPTPRGRRTTKGWGRSPAPFGAARA RRPPTSGQHSAMAASSGG
Frame 3: RSGLGPEHVPDPRV

Colon Tumour cDNA Library – Round 9 – Plaque D1

GGTTTCGAGGGAC**GAATTCA**AGCCTTCCGCTTCGGACCGAGGGCAGTAGGCTCTCGGCTCCTGGTCCCACTGC
TGCTCAGCCCAGTGGCCTCACAGGACACCAGCTTCCCAGGAGGCGTCTGACACAGTATGATGATGAAGATCCC
ATGGGGCAGCATCCCAGTACTGATGTTGCTCCTGCTCCTGGGCCTAATCGATATCTCCCAGGCCCAGCTCAGC
TGCACCGGGCCCCCAGCCATCCCTGGCATCCCGGGTATCCCTGGGACACCTGGCCCCGATGGCCAACCTGGGA
CCCCAGGGATAAAAGGAGAGAAAGGGCTTCCAGGGCTGGCTGGAGACCATGGTGAGTTCGGAGAGAAGGGAGA
CCCAGGGATTCCTGGGAATCCAGGAAAAGTCGGCCCCAAGGGCCCCATGGGCCCTAAAGGTGGCCCAGGGGCC
CCTGGAGCCCCAGGCCCCAAAGGTGAATCGGGAGACTACAAGGCCACCCAGAAAATCGCCTTCTCTGCCACAA
GAACCATCAACGTCCCCCTGCGCCGGGACCAGACCATCCGCTTCGACCACGTGATCACCAACATGAACAACAA
TTATGAGCCCCGCAGTGGCAAGTTCACCTGCAAGGTGCCCGGTCTCTACTACTTCACCTACCACGCCAGCTCT
CGAGGGAACCTGTGCGTGAACCTCATGCGTGGCCGGGAGCGTGCACAGAAGGTGGTCACCTTCTGTGACTATG
CCTACAACACCTTCCAGGTCACCACCGGTGGCATGGTCCTCAAGCTGGAGCAGGGGGAGAACGTCTTCCTGCA
GGCCACCGACAAGAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTCTTTG
GGGGGGATAA (up primer)

LOCUS NM_000491 1044 bp mRNA linear PRI 24-SEP-2011
DEFINITION **Homo sapiens complement component 1, q subcomponent, B chain (C1QB), mRNA.**
ACCESSION NM_000491 VERSION NM_000491.3 GI:87298827

1 gcccttcccg cctctgggga agggaacttc cgcttcggac cgagggcagt aggctctcgg
61 ctcctggtcc cactgctgct cagcccagtg gcctcacagg acaccagctt cccaggaggc
121 gtctgacaca gt**atgatgat gaagatccca tggggcagca tcccagtact gatgttgctc**
181 **ctgctcctgg gcctaatcga tatctcccag gcccagctca gctgcaccgg gcccccagcc**
241 **atccctggca tcccgggtat ccctgggaca cctggccccg atggccaacc tgggacccca**
301 **gggataaaag gagagaaagg gcttccaggg ctggctggag accatggtga gttcggagag**
361 **aagggagacc cagggattcc tgggaatcca ggaaaagtcg gccccaaggg ccccatgggc**
421 **cctaaaggtg gcccaggggc ccctggagcc ccaggcccca aaggtgaatc gggagactac**
481 **aaggccaccc agaaaatcgc cttctctgcc acaagaacca tcaacgtccc cctgcgccgg**
541 **gaccagacca tccgcttcga ccacgtgatc accaacatga acaacaatta tgagccccgc**
601 **agtggcaagt tcacctgcaa ggtgcccggt ctctactact tcacctacca cgccagctct**
661 **cgagggaacc tgtgcgtgaa cctcatgcgt ggccgggagc gtgcacagaa ggtggtcacc**
721 **ttctgtgact atgcctacaa caccttccag gtcaccaccg gtggcatggt cctcaagctg**
781 **gagcaggggg agaacgtctt cctgcaggcc accgacaaga actcactact gggcatggag**
841 **ggtgccaaca gcatcttttc cgggttcctg ctctttccag atatggaggc ctga**cctgtg
901 ggctgcttca catccacccc ggctcccccc gccagcaacg ctcactctac ccccaacacc
961 acccccttgcc caaccaatgc acacagtagg gcttggtgaa tgctgctgag tgaatgagta
1021 aataaactct tcaaggccaa ggga //

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame):

Frame 1: **SLPLRTEGSRLSAPGPTAAQPSGLTGHQLPRRRLTQYDDEDPMGQHPSTDVAPAPG**
PNRYLPGPAQLHRAPSHPWHPGYPWDTWPRWPTWDPRDKRRERASRAGWRPW
(does not match above gene; blastp brings low similarity hit with *capsaspora* sp.)
Frame 2: AFRFGPRAVGSRLLVPLLLSPVASQDTSFPGGV
(after this aa sequence is a stop codon, then the protein sequence matching the above gene comes up)
Frame 3: PSASDRGQ

Colon Tumour cDNA Library – Round 9 – Plaque E2

GTTGTAGATTGGGCCGAGAC**GAATTCC**ACAGAGAAAAACATGATTAGCATATATCAGAAAGCAATTATCTATT
CTACATGATGTAGTAATAATGACAGTAATGAATAGCTAACACATAACATTTACTGTGTGCCAAGACACTGTTG
CAAATGACTTTCATATATAAATTCATTGAATCTCCACAACAATCCTTTGAAGTGGGTAAACTATTATTATGTC
ATTTTAACAGACAGAGGAACTGAGGTATAGAGAGGTTTAGAAACTTATCCCGGGTTACAGAGCTAGTAAGGGG
CTGAACCTAGACTTGATCTTAGAACACGTGGATCTAGAGTCTAAGCTATGAAGCTTGCGGCCGCACTCGAGTA
ACTAGTTAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTAAAACTAGTTACTCGAGTGCGCGCAAGCTT
CATAGCTTAGACTCTAGATCCACGTGTTCTAAGATCAAGTCTAGGTTCAGCCCCTTACTAGCTCTGTAACCCG
GGATAAGTTTCTAAACCTCTCTATACCTCAGTTCCTCTGTCTGTTAAAATGACATAATAATAGTTTACCACTT
CAAAGGATTGTTGTGGAGATTAATGAATTTATATATGAAAGTATTTCACAGTGTCTTGGCACACAGTAAATGT
TATGTGTTGCTATTATTACTGTCATTATTACTACATCAGTAGAATAGATAATTGCTTTCTGATATATGCTAAT
CATGTTTTTTCTCTGTGGAATTCGGATCCCGAGCATCACTCAGGGGAAAACACAGCCCCAAAA (up
primer)

CCAAGACTTCTCGAGTGCGGCCGCAAGCTTCATAGCTTAACTCTAGATCCACGTGTTCTAAGATCAAGTCTAG
GTTCAGCCCCTTACTAGCTCTGTAACCCGGGATAAGTTTCTAAACCTCTCTATACCTCAGTTCCTCTGTCTGT
TAAAATGACATAATAATAGTTTACCCACTTCAAAGGATTGTTGTGGAGATTCAATGAATTTATATATGAAAGT
CATTTGCAACAGTGTCTTGGCACACAGTAAATGTTATGTGTTAGCTATTCATTACTGTCATTATTACTACATC
ATGTAGAATAGATAATTGCTTTCTGATATATGCTAATCATGTTTTTCTCTGTGGAATTCGGATCCCCGAGCAT
CACACCTGACTGGAATACGACAGCTCCAAA (down primer)

LOCUS AC084083 209859 bp DNA linear PRI 30-AUG-2002
DEFINITION **Homo sapiens chromosome 8, clone RP11-550A5, complete sequence.**
ACCESSION AC084083 VERSION AC084083.10 GI:22549881

no CDS
match from position 185196-185519

Peptide encoded by Phage DNA:

Frame 1: STEKNMISIYQKAIIYST
Frame 2: PQRKT
Frame 3: HREKHD

Colon Tumour cDNA Library – Round 9 – Plaque H2

CCAGCCCCCTCGGTCTGGAC**GAATTCT**GTTCTCCCAGAGCTGCACCTGCCTCGCAGAGGCCAGCACCCCACTC
TCCTGCCTCCAGTGGCCCTGCCGCAGATGTCTCCCAAAAAGTTGAGCCTTTCTAGATGGCTTAGGTGGCACCA
TGGCTCAGCAGGAGGGGCGGGAGGCACCAGGGTTCTTGTTTGGACCCTGCCCCTGGGCCATGGCCAGGTGACC
ATGGCTACATTGCCAAACCTCTGACTGCCACAGCTGCAGACTGAGAGGGTGGGTCTGAGTCCCCACAATGTCT
GAAGCTGCCCCTGGGATTCTCAGGCCAACCTGCCAACAGCAAGCGGATTTTCTTGCAAGATCAGGGACCCCAT
TTCTGCAGCCAGTGTCTCCTGGGTGCCTTCTGAGGACTCCCACCCCCATCCCAGTATCTCATCTGTCCCCTCT
CCTGGGGCTTAAGTGGGTTGCTTCCAGGCAGAAGCAGCCAAGGACCGATTCCAGGCACTTTCTGTAGCAAATG
ACTGTGAATTACGACTTCTCTTGCCCTTCTTCTAGCAGTCTGTGCCTCCTCTCTGACCAGTTTGGAGGGCACT
GAAGAAAGGCAAGGGCCGTGCTGCTGCTGGGCGGGGCAGGAGAGGAGCCTGGCCAGTGTGCCACATTAAATAC
CCGTGCAGGCGCGGAGAAGCAACCGGCACCCCCTTCCGGCCTGAAAGCCCTCCCTGCAAGAAGGTGTGCAGGA
GAGAAGAGGCCCCGGCATGGGGATCTGGGTTCTAGAGGGCATGTGATGACTGTAAATGTTCACTGGGTGGGTA
GGGAGTGGTATCCAGTGTTCAAGTGCAGAAATCTTTGGCTTTGCTACCAGTTCCATATGATGAGAAATAAACG
TTCGCTGAGGTTTTGTTTCATAACAAAAAAGAAAAAAAAAAAAAAAAAAAAAGGTTTTGGCGCCCCCCCACAA
AATAATTTACCCCTGGGGCCCTCTA (up primer)

CGGGGCCACTGACTCTCCAGTGCGGCCGCAGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCCA
AAAAAAACCCCCCAAAAGTTTTTTTTTTCCCCAATTGGAAGGGGTGGCTTAGGTGGAATTTTTCCTCTTTAGA
GGGGCGGCCCCCCCCCCCCCCCCCGTTTGAACTTTGCCCCTGGGCCATGGCCCCGAGAACACGGAAACCTTGC
CGGGGCTCTTTTTGCCCCCGCTGCCTTTTGTGAGGGGGGGTCTTATTCCCCGCGGGGGGGGGGCGGCCTTTC
CCCCCCCCCGGCCGGGTTTTTAATGGGGCCCGGATTTTCTTGCCCCCCCCCCCCCCCTTTCTGCAGCCCGGC
CCTTTTTTTTTTTTTCTGACCCCTCCCACCCCCAAAAAAGGGGTCAACTGTCCCCAAAAAGGGGAAAAAAATGG
GTTGCTTCCAGGCTTATTCAACCAAAGACCGATTCCAGGCACTTTCTGTAGCTTATGACTGGGAATTACCACT

TCTCTCCCCCTTCTGGGGGCAAAATGAGACTCCTGGGGGAGGGGGGGGGAGGGCACTGAAGAAAGGCAAGGAC
CGTGCTGCTGCTGGGGGGGGCAGGAGAGGAGCCAGAACACTGTGCCACATTAAATACCCGTGCAGGAACGGAG
AAGCAACCTGCACCCCCTTCGGGCCTGAAAGCCCTCCCTGCAAGAAGGTGTGCAGGAGAGAAGAGGCCCCGGC
ATGGGGATCTGGGTTCTAGAGGGCATGTGATGACTGTAAATGTTCACTGGGTGGGTAGGGAGTGGCATCCACT
GTTCAAGTGCAGAGATCCTTTGGCTTTGCTACCAGTTCCATATGATTTATAAATAAACGTTCGCTGAGTTTTG
TTTCGTACAAAAAAAAGGGAGGTTAGCCTTATAAAAGAGAGAGTCGTTCTCGCGAAACAAAAAAATTTAATCC
CCCCCTTTGCGCCCCTAAAGGGGATTGAGGGGTAAAAAATTCTTGTTTCTGTGGGAGGCGGCAGCAGTTTTTT
TTTTTTCTCCTATGTTTTTCTTGTTTTGTTGATACA (down primer)

LOCUS AY123223 3535 bp mRNA linear PRI 06-MAY-2003
DEFINITION **Homo sapiens sestrin 2 mRNA, complete cds.**
ACCESSION AY123223 VERSION AY123223.1 GI:22900836

CDS 323-1765
1 ctccgcggcg gggatgctga ggagcgctgg gtccgggagc agccctggcc cctgcggact
61 tccgaggccg tgaaaacccc tgcgctgcgg cccttcccag gcccccgagg ccgttcgccg
121 ttcccgaagc ccgactgggg gaagagtcca gcaccaaagc ggccgttctc ggattccgga
181 gcgttctgga gccccgagag acgccccggg gttctagaag ctccccggcg gcgcccagtc
241 ccggcttcat tcgggcgtcc ctccgaaacc cactcgggtg cacgggtcgt cggcgagccg
301 cgaccgggtc ctggcgcgca cc**atgatcgt ggcggactcc gagtgccgcg cagagctcaa**
361 **ggactacctg cggttcgccc cgggcggcgt cggcgactcg ggccccggag aggagcagag**
421 **ggagagccgg gctcggcgag gccctcgagg gcccagcgcc ttcatccccg tggaggaggt**
//
1621 **gaagaccacc cgaagaatgt acaacctctt ctggaggcac ttccgccact cagagaaggt**
1681 **ccacgtgaac ttgctgctcc tggaggcgcg catgcaagcc gctctgctgt acgccctccg**
1741 **tgccatcacc cgctacatga cctga**ctcct gagcaggacc tgggcccggt tcagctcccc
1801 acaaggactt ctctgtctgg agacagcccc agacccttt gtgtcccatg cccaccctcc
//
2521 ctttatgctt gaggttccaa cctggagcca cagtgtgtga naggaggagg agagggagaa
2581 ttctgttctc ccanagctgc acctgcctcg canaggccag caccccactc tcctgcctcc
2641 agtggccctg ccgcagatgt ctcccaaaaa gttgagcctt tctagatggc ttaggtggca
//
3301 ggcatgggga tctgggttct agagggcatg tgatgactgt aaatgttcac tgggtgggta
3361 gggagtggta tccagtgttc aagtgcagaa atctttggct ttgctaccag ttccatatga
3421 tgagaaataa acgttcgctg aggttttgtt tcataaaaaa aaaaaaaaaa aaaaaaaaaa
3481 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaa

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame):

Frame 1: SVLPELHLPRRGQHPTLLPPVALPQMSPKKLSLSRWLRWHHGSAGGAGGTRVLVW
TLPLGHGQVTMATLPNL
Frame 2: LFSQSCTCLAEASTPLSCLQWPCRRCLPKS
Frame 3: CSPRAAPASQRPAPHSPASSGPAADVSQKVEPF

8.3.3.4.4 Human liver tumour

Liver Tumour cDNA Library – Round 9 – Plaque B4

CCATGCATTCGTCGCGAC**GAATTCA**GCCGAGATCGGGCTTGGGCCCAGAGCATGTTCCAGATCCCAGAGTTTG
AGCCGAGTGAGCAGGAAGACTCCAGCTCTGCAGAGAGGGGCCTGGGCCCCAGCCCCGCAGGGGACGGGCCCTC
AGGCTCCGGCAAGCATCATCGCCAGGCCCCAGGCCTCCTGTGGGACGCCAGTCACCAGCAGGAGCAGCCAACC
AGCAGCAGCCATCATGGAGGCGCTGGGGCTGTGGAGATCCGGAGTCGCCACAGCTCCTACCCCGCGGGGACGG
AGGACGACGAAGGGATGGGGGAGGAGCCCAGCCCCTTTCGGGGCCGCTCGCGCTCGGCGCCCCCCAACCTCTG
GGCAGCACAGCGCTATGGCCGCGAGCTCCGGAGGATGAGTGACGAGTTTGTGGACTCCTTTAAGAAGGGACTT
CCTCGCCCGAAGAGCGCGGGCACAGCAACGCAGATGCGGCAAAGCTCCAGCTGGACGCGAGTCTTCCAGTCCT
GGTGGGATCGGAACTTGGGCAGGGGAAGCTCCGCCCCCTCCCAGTGACCTTCGCTCCACATCCCGAAACTCCA
CCCGTTCCCACTGCCCTGGGCAGCCATCTTGAATATGGGCGGAAGTACTTCCCTCAGGCCTATGCAAAAAGAG

GATCCGTGCTGTCTCCTTTGGAGGGAGGGCTGACCCAGATTCCCTTCCGGTGCGTGTGAAGCCACGGAAGGCT
TGGTCCCATCGGAAGTTTTGGGTTTTCCGCCCACAGCCGCCGGAAGTGGCTCCGTGGCCCCGCCCTCAGGCTC
CGGGCTTTCCCCCAGGCGCCTGCGCTAAGTCGCGAGCCATGTTTAACCGTTGCGTCACCGGGACCCGAGCCCC
CGCGATGCCCTGGGGGCCGTGCTCACTACCAAATGTTAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCC
TTGGGGCCTCTAAACGGGCTCTTGGAAGGGGTTAACA (up primer)

GGGTTCTTAAGAGACTTATCCAGTGCAGACGCGGCTGGCATTTGGTATGGACCACAGCCCCCAGGTTGTCCCG
AGGGCTCGGGTCACTCCGACTCTGCGGATAGGGGCCTGTCCCCACTTCCCGCAGGGGACTGGGCCAAAGGCTC
GAGCCTGAGGGCGGGGCCACCCCCAGGCTTCCTGTGGCTGTGGGTCGACAACCCAAAACTTCAAATGGGACCA
ACCATTCTGTGGCTTCTGGGGCTGTGGAGATCCTGAGGGTCACCCTCCTACCCCGGAGACACCGAGGATCATC
ATTTTATGGGGGAGGAGCGGAAGTCCTTTCGGGGATATTCGAGATGGGTGCCCCCGGACCTGTGAACGGGTGG
AGTTATGGGATGTGGAGCCGGAGGACTGGGACGAGGTTGAGCTTCCCTTGCCCAAGGGACTATCCCACCCGAC
TGGAAGACTCCCGTCCCGCTGATGCTTTGCAGCTCCTGCTGGACTGTAGCCTTCCTCTTCTGGTGAGAACGTA
CCTTGTTAAGGGAAAGCTCCGCCCCCTCACTCTGACTTTCGATCTCCCTCCCATAACTCTGTCCTGTCCCAAT
GTTGGGGGGCGCCATCTTGAATATGGGCGGAAAGCTTCCCTCTCCTCCTCCATCCCTTCGATCCCCTCTGTCT
CCTTTGGAAGGAGGTGTGACCCACTTTCGATCTCCGGTGCGTGTGAAGCTCCATGATGCTTGCTGCTGGTTGA
CTGTTCCTGTTTTCCGACTGGAGTCCCACAGAAGTGCCTCCGTGCTGCCGATCATGCGCTCCGGACTCTGACC
CCCGGTCCTGTGCTAGGTTGGGAGCCAGGCTTCACTCTTGCGACATGGAGACTTCATGCCCCCTGATCTCATG
CTGGGGCGAGCTGGATAATGAATGTGAAGCTTGCCCCCGCTCGCCAGGAATTCGTATCCCCCATGA (down
primer)

LOCUS NM_032989 986 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens BCL2-associated agonist of cell death (BAD), transcript variant 2, mRNA.**
ACCESSION NM_032989 VERSION NM_032989.2 GI:197116382

```
  1 aactagggcc cggagcccgg ggtgctggag ggaggcggca ggcccgggtc aggggcctcg
 61 agatcgggct tgggcccaga gcatgttcca gatcccagag tttgagccga gtgagcagga
121 agactccagc tctgcagaga ggggcctggg ccccagcccc gcaggggacg ggccctcagg
181 ctccggcaag catcatcgcc aggccccagg cctcctgtgg gacgccagtc accagcagga
241 gcagccaacc agcagcagcc atcatggagg cgctggggct gtggagatcc ggagtcgcca
301 cagctcctac cccgcgggga cggaggacga cgaagggatg ggggaggagc ccagcccctt
361 tcggggccgc tcgcgctcgg cgccccccaa cctctgggca gcacagcgct atggccgcga
421 gctccggagg atgagtgacg agtttgtgga ctcctttaag aagggacttc ctcgcccgaa
481 gagcgcgggc acagcaacgc agatgcggca aagctccagc tggacgcgag tcttccagtc
541 ctggtgggat cggaacttgg gcaggggaag ctccgccccc tcccagtgac cttcgctcca
601 catcccgaaa ctccacccgt tcccactgcc ctgggcagcc atcttgaata tgggcggaag
661 tacttccctc aggcctatgc aaaaagagga tccgtgctgt ctcctttgga gggagggctg
721 acccagattc ccttccggtg cgtgtgaagc cacggaaggc ttggtcccat cggaagtttt
781 gggttttccg cccacagccg ccggaagtgg ctccgtggcc ccgccctcag gctccgggct
841 ttcccccagg cgcctgcgct aagtcgcgag ccaggtttaa ccgttgcgtc accgggaccc
901 gagccccccgc gatgccctgg gggccgtgct cactaccaaa tgttaataaa gcccgcgtct
961 gtgccgccga aaaaaaaaaa aaaaaa //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined complete CDS):

```
Frame 1: EFSRDRAWAQSMFQIPEFEPSEQEDSSSAERGLGPSPAGDGPSGSGKHHRQAPGLLWDAS
         HQQEQPTSSSHHGGAGAVEIRSRHSSYPAGTEDDEGMGEEPSPFRGRSRSAPPNLWAAQR
         YGRELRRMSDEFVDSFKKGLPRPKSAGTATQMRQSSSWTRVFQSWWDRNLGRGSSAPSQ
Frame 2: NSAEIGLGPRACSRSQSLSRVSRKTPALQRGAWAPAPQGTGPQAPASIIARPQASCGTPV
         TSRSSQPAAAIMEALGLWRSGVATAPTPRGRRTTKGWGRSPAPFGAARARRPPTSGQHSA
         MAASSGG
Frame 3: IQPRSGLGPEHVPDPRV
```

Liver Tumour cDNA Library – Round 9 – Plaque C4

GGGGTCGGCTAC**GAATTCA**TGGCTGCACTACACCGATTCTGAAAAGGAGGAACAAACGCCCCGTTTGGAACAG
AATCTGGCCCATCCCATGGAGAGGGCCACTGTGCCTGACAGATTTGTTGGGGTGAATCTGGGTGCCCAGGGAT
CAATATGGCAACTCTACACTCTTGCTGGCGAGGGAACTGGAGGCCAGCACCAAGGAGACCCAAACAGGCCTCC
CAGCCAGGCACAGCCCAGGGCACCCTGGGTTAGCCTGATCCCGCCCTGGTGCCTGGGTCCTTCACGCGGCTCT
CCTGTAGACTTTCTATTAAGAAATGCCTGTGGTAGAGTCAGGCGGGTGAGAAGTCTACACTCCTTGGCAGTAC
CTTGAAAGGAAAAATCGTCACAGATAAAAAAAAATAAACAGGTGAGAAATATCAACGCAGCAGCAGCAGTGTT
GAAAGGGACACAGCTTGAGGGGTGGATCTGGGGGTGAAATTCAGGTTAAGACGGAGTGCAGGGTGCTAGAAGC
TGCCACGCAGGTTGATGAGCCACGTGCTATGAAAGCCAGGAGCCACTGACCCGCTCAGCTCTTGGTTTCCTTC
AGCCGCCAGCTGGGAGGCTCTTGCCCCACCGCCCTAGGGACCCAAGCTATCTAGAAGAGGAGCCAGCCGTCTC
TGCGCAAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTTTGGAAGGGGGT
TAAA (up primer)

BLASTn search strategies:

- No similarity via RNA database: "others, Reference RNA Sequences (refseq_rna)", therefore extend search: outside *Homo sapiens* also results in no hit; more dissimilar sequences brings no hit.
- "others, refseq_dna" results in 4 human hits, all chromosome 22 genomic contig, featuring protein MICAL-3 isoform 1, but "gene in backwards"
- "others, nucleotide collection nr/nt, *Homo sapiens*": chromosome 22, plus/plus, but gDNA.

Liver Tumour cDNA Library – Round 9 – Plaque B5

GAAATCTTATAC**GAATTCA**GCATACATCTCACACACAGGAGGCACTGGTATTTAGAAACTTAGGAAAATTACT
AGAATTGTTAAGACTGAGGTGAAGAATTTGTACAGCCTTCCACAATTTCAAAGAGGTAGTGATAGATACTTGT
CCTCCTGGCAATATCAAATGCAGTTTCTTCCAAGTTGTTTTTCAGCCCTGGTTTGACGTAACGGTTCATCAGG
AGGAGTTCTAGGGTATCCTTGCTGTCTCTGTTCCCAGCAGCAAGATGCAAGGGGGTCAAGAGGCCTTTTGTTT
GGGCATTGATATCTGCATCATGCTGCAGTAAGAAAGAAGCCACTCTGGTATTATTCCACTTACAAGCACTGTG
CAGGGGCGTCCAGCCATCCACAGTCACTGCATGAACATCGGCCCCCTGTGCAATGAGCTCCTGAACAATATCT
AAGTGTCCACTGTAGGCTGCTCGATGAAGAGGGGTATACTCATCTTCATCCCTAGTGTTCACGTGAGTGGCCT
TTTCAGAAAGGAGTCTCCGCACTGTGGTAAGAATCAAGAGATAGGTATGATGGCCATAACAGTAAAGGCAGGT
CTTCTGGCTTGAACCCAGTGCTTTTCCTACTGTCACACTTCCCTCTGTAATTTTTTTTTTTACCCGATTTTTT
TCAGCAGCCCAAAGAAGCAATCTGCTTGGGTCTTTTTCCATTTTTTTTCTTGCAAAGCTTGCGGCCGCACTCG
AGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTCTTTGAGGGGGTTATTAAACTAGTTACTCGAGTGC
GGCGCAAGCTTTGC (up primer)

CGGGGCCTACGAGATTCTCGAGTGCGGCCGCAAGCTTTGCAAGAAAAAAAAATGGAAAAAGACCCAAGCAGATT
GCTTCTTTGGGCTGCTGAAAAAAATCGGGTAAAAAAAAAAAATTACAGAGGGAAGTGTGACAGTAGGAAAAGCA
CTGGGTTCAAGCCAGAAAACCTGCCTTTACTGTTATGGCCATCATACCTATCTCTTGATTCTTACCACAGTGC
GGAAACTCCTTTCTGAAAAGGCCACTCACGTGAACACTAGGGATGAAGATGAGTATACCCCTCTTCATCGAGC
AGCCTACAGTGGACACTTAGATATTGTTCAGGAGCTCATTGCACAGGGGGCCGATGTTCATGCAGTGACTGTG
GATGGCGGGACGCCCCTGCACAGTGCTTGTAAGTGGAATAATACCCCAGTGGCTTCTTTCTTACTGCACCATG
ATGCAGATATCAATGCCCAAACAAAAGGCCTCTTGACCCCCTTGCTTCTTGCTGCTGGGAACAAAGACAGCGG
GGATACCCTAAAACTCCTCCTGATGAACCGTTACATCAAACCAGGGCTGAAAAACAACTTGAAAGAAACTGCA
TTTGATATTGCCAGGAGGACAAGTATCTATCACTACCTCTTTGAAATTGGGGAAGGCTGTACAAATTCTTCAC
CTCAGTCTTAACAATTCTAGTAATTTTCCTAAGTTTCTAAATACCATTGCCTCCGGTGTGTGAGATGTATTGC
TTGAATTCGGATCCCCGACCATCACACCTGACTGGAATACGACGGCGTCCATA (down primer)

LOCUS NG 007261 83575 bp DNA linear PRI 27-MAY-2012
DEFINITION **Homo sapiens MRE11 meiotic recombination 11 homolog A (S. cerevisiae) (MRE11A), RefSeqGene (LRG 85) on chromosome 11.**
ACCESSION NG_007261 VERSION NG_007261.1 GI:163965385

CDS multiple positions starting at 6074

```
   1 caacaaagat ccaggatgtg ggatatttaa aaatacattt ctcttgtatc agcaccaaaa
  61 ataccagttt ctgtacaacc cgtgtcaagt ttttaagtat acacacatta aaattatttc
 121 tgataaaaaa atcacaaact ccacagaaat aattaaagat atatgcaaat tactttcaaa
```

```
 181 tacattttga catcaaagtc aggcaattca cttggaagaa tgttataatg aatgtcactg
 241 aatttttgta gtaagatgtt tgacgtcaac tttgattatg ggaatacatc tcacacacag
 301 gaggcactgg tatttagaaa cttaggaaaa ttactagaat tgttaagact gaggtgaaga
 361 atttgtacag ccttccacaa tttcaaagag gtagtgatag atacttgtcc tcctggcaat
 421 atcaaatgca gtttcttcca agttgttttt cagccctggt ttgacgtaac ggttcatcag
 481 gaggagttct agggtatcct tgctgtctct gttcccagca gcaagatgca agggggtcaa
 541 gaggcctttt gtttgggcat tgatatctgc atcatgctgc agtaagaaag aagccactct
 601 ggtattattc cacttacaag cactgtgcag gggcgtccag ccatccacag tcactgcatg
 661 aacatcggcc ccctgtgcaa tgagctcctg aacaatatct aagtgtccac tgtaggctgc
 721 tcgatgaaga ggggtatact catcttcatc cctagtgttc acgtgagtgg ccttttcaga
 781 aaggagtctc cgcactgtgg taagctgaga agaaaaaaaa gaaatctttc aactcaactc
 841 aaaccaaaac aataactaaa acaaaattgt cttgtggtgt tagatttaat taagttgaat
 901 aaaaacttac ttgttagtag atgactttaa aatgaaaata atttaagatc tcttttgggg
  //
83521 tcgtgtttat attttatgta aaaatcagaa ttttctacat agaagcttat aactg
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame):

Frame 1: SAYISHTGGTGI
Frame 2: QHTSHTQEALVFRNLGKLLELLRLR
Frame 3: SIHLTHRRHWYLET

Liver Tumour cDNA Library – Round 9 – Plaque C5

TGGGGGGTTCGGGAC**GAATTCA**AGCAATACATCTCACACACAGGAGGCACTGGTATTTAGAAACTTAGGAAAA
TTACTAGAATTGTTAAGACTGAGGTGAAGAATTTGTACAGCCTTCCACAATTTCAAAGAGGTAGTGATAGATA
CTTGTCCTCCTGGCAATATCAAATGCAGTTTCTTCCAAGTTGTTTTTCAGCCCTGGTTTGACGTAACGGTTCA
TCAGGAGGAGTTCTAGGGTATCCTTGCTGTCTCTGTTCCCAGCAGCAAGATGCAAGGGGGTCAAGAGGCCTTT
TGTTTGGGCATTGATATCTGCATCATGCTGCAGTAAGAAAGAAGCCACTCTGGTATTATTCCACTTACAAGCA
CTGTGCAGGGGCGTCCAGCCATCCACAGTCACTGCATGAACATCGGCCCCCTGTGCAATGAGCTCCTGAACAA
TATCTAAGTGTCCACTGTAGGCTGCTCGATGAAGAGGGGTATACTCATCTTCATCCCTAGTGTTCACGTGAGT
GGCCTTTTCAGAAAGGAGTCTCCGCACTGTGGTAAGAATCAAGAGATAGGTATGATGGCCATAACAGTAAAGG
CAGGTCTTCTGGCTTGAACCCAGTGCTTTTCCTACTGTCACACTTCCCTCTGTAATTTTTTTTTTTACCCGAT
TTTTTTCAGCAGCCCAAAGAAGCAATCTGCTTGGGTCTTTTTCCATTTTTTTTCTTGCAAAGCTTGCGGCCGC
ACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGCTTGGAGGGGGTTA (up primer)

LOCUS NM_017704 1924 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens ankyrin repeat domain 49 (ANKRD49), mRNA.**
ACCESSION NM_017704 VERSION NM_017704.2 GI:41350197

Gene in backwards

Liver Tumour cDNA Library – Round 9 – Plaque C6

CGTTAG**GAATTCC**GGGGCTTGGGCCCAGAGCATGTTCCAGATCCCAGAGTTTGAGCCGAGTGAGCAGGAAGAC
TCCAGCTCTGCAGAGAGGGGCCTGGGCCCCAGCCCCGCAGGGGACGGGCCCTCAGGCTCCGGCAAGCATCATC
GCCAGGCCCCAGGCCTCCTGTGGGACGCCAGTCACCAGCAGGAGCAGCCAACCAGCAGCAGCCATCATGGAGG
CGCTGGGGCTGTGGAGATCCGGAGTCGCCACAGCTCCTACCCCGCGGGGACGGAGGACGACGAAGGGATGGGG
GAGGAGCCCAGCCCCTTTCGGGGCCGCTCGCGCTCGGCGCCCCCCAACCTCTGGGCAGCACAGCGCTATGGCC
GCGAGCTCCGGAGGATGAGTGACGAGTTTGTGGACTCCTTTAAGAAGGGACTTCCTCGCCCGAAGAGCGCGGG
CACAGCAACGCAGATGCGGCAAAGCTCCAGCTGGACGCGAGTCTTCCAGTCCTGGTGGGATCGGAACTTGGGC
AGGGGAAGCTCCGCCCCCTCCCAGTGACCTTCGCTCCACATCCCGAAACTCCACCCGTTCCCACTGCCCTGGG
CAGCCATCTTGAATATGGGCGGAAGTACTTCCCTCAGGCCTATGCAAAAAGAGGATCCGTGCTGTCTCCTTTG
GAGGGAGGGCTGACCCAGATTCCCTTCCGGTGCGTGTGAAGCCACGGAAGGCTTGGTCCCATCGGAAGTTTTG
GGTTTTCCGCCCACAGCCGCCGGAAGTGGCTCCGTGGCCCCGCCCTCAGGCTCCGGGCTTTCCCCCACGCGCC
TGCGCTAAGTCGCGAGCCAGGTTTAACCGTTGCGTCACCGGGACCCGAGCCCCGCGATGCCCTGGGGGCCGTG

CTCACTACCAAATGTTAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAACGCCTTTG
TGAGAGAAGGGTTT (up primer)

LOCUS NM_032989 986 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens BCL2-associated agonist of cell death (BAD), transcript variant 2, mRNA.**
ACCESSION NM_032989 VERSION NM_032989.2 GI:197116382

```
  1 aactagggcc cggagcccgg ggtgctggag ggaggcggca ggcccgggtc aggggcctcg
 61 agatcgggct tgggcccaga gcatgttcca gatcccagag tttgagccga gtgagcagga
121 agactccagc tctgcagaga ggggcctggg ccccagcccc gcaggggacg ggccctcagg
181 ctccggcaag catcatcgcc aggccccagg cctcctgtgg gacgccagtc accagcagga
241 gcagccaacc agcagcagcc atcatggagg cgctggggct gtggagatcc ggagtcgcca
301 cagctcctac cccgcgggga cggaggacga cgaagggatg ggggaggagc ccagcccctt
361 tcggggccgc tcgcgctcgg cgccccccaa cctctgggca gcacagcgct atggccgcga
421 gctccggagg atgagtgacg agtttgtgga ctcctttaag aagggacttc ctcgcccgaa
481 gagcgcgggc acagcaacgc agatgcggca aagctccagc tggacgcgag tcttccagtc
541 ctggtgggat cggaacttgg gcaggggaag ctccgccccc tcccagtgac cttcgctcca
601 catcccgaaa ctccacccgt tcccactgcc ctgggcagcc atcttgaata tgggcggaag
661 tacttccctc aggcctatgc aaaaagagga tccgtgctgt ctcctttgga gggagggctg
721 acccagattc ccttccggtg cgtgtgaagc cacggaaggc ttggtcccat cggaagtttt
781 gggttttccg cccacagccg ccggaagtgg ctccgtggcc ccgccctcag gctccgggct
841 ttcccccagg cgcctgcgct aagtcgcgag ccaggtttaa ccgttgcgtc accgggaccc
901 gagcccccgc gatgccctgg gggccgtgct cactaccaaa tgttaataaa gcccgcgtct
961 gtgccgccga aaaaaaaaaa aaaaaa //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = complete translation of CDS):

Frame 1: **GAWAQSMFQIPEFEPSEQEDSSSAERGLGPSPAGDGPSGSGKHHRQAPGLLWD ASHQQEQPTSSSHHGGAGAVEIRSRHSSYPAGTEDDEGMGEEPSPFRGRSRSA PPNLWAAQRYGRELRRMSDEFVDSFKKGLPRPKSAGTATQMRQSSSWTRV FQSWWDRNLGRGSSAPSQ**
Frame 2: GLGPRACSRSQSLSRVSRKTPALQRGAWAPAPQGTGPQAPASIIARPQASCGT PVTSRSSQPAAAIMEALGLWRSGVATAPTPRGRRTTKGWGRSPAPFGAARARR PPTSGQHSAMAASSGG
Frame 3: GLGPEHVPDPRV

Liver Tumour cDNA Library – Round 9 – Plaque G6

CAGCTACGTACACC**GAATTCT**GAAAAGGAGGAACAAACGCCCCGTTTGGAACAGAATCTGGCCCATCCCATGG
AGAGGGCCACTGTGCCTGACAGATTTGTTGGGGTGAATCTGGGTGCCCAGGGATCAATATGGCAACTCTACAC
TCTTGCTGGCGAGGGAACTGGAGGCCAGCACCAAGGAGACCCAAACAGGCCTCCCAGCCAGGCACAGCCCAGG
GCACCCTGGGTTAGCCTGATCCCGCCCTGGTGCCTGGGTCCTTCACGCGGCTCTCCTGTAGACTTTCTATTAA
GAAATGCCTGTGGTAGAGTCAGGCGGGTGAGAAGTCTACACTCCTTGGCAGTACCTTGAAAGGAAAAATCGTC
ACAGATAAAAAAAAATAAACAGGTGAGAAATATCAACGCAGCAGCAGCAGTGTTGAAAGGGACACAGCTTGAG
GGGTGGATCTGGGGGTGAAATTCAGGTTAAGACGGAGTGCAGGGTGCTAGAAGCTGCCACGCAGGTTGATGAG
CCACGTGCTATGAAAGCCAGGAGCCACTGACCCGCTCAGCTCTTGGTTTCCTTCAGCCGCCAGCTGGGAGGCT
CTTGCCCCACCGCCCTAGGGACCCAAGCTATCTAGAAGAGGAGCCAGCCGTCTCTGCGCAAAGCTTGCGGCCG
CACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGTTTGTAGGGGGGGTAAAAA (up primer)

BLAST search strategies:
- No similarity via RNA database: "others, Reference RNA Sequences (refseq_rna)", therefore extend search: outside *Homo sapiens* also results in no hit; more dissimilar sequences brings no hit.

- "others, refseq_dna" results in 4 human hits, all chromosome 22 genomic contig, featuring protein MICAL-3 isoform 1, but gene in backwards
- "others, nucleotide collection nr/nt, *Homo sapiens*": chromosome 22, plus/plus, but gDNA
- Compare to plaque "Lung Tumour F7"

8.3.3.4.5 Human lung tumour

Lung Tumour cDNA Library – Round 9 – Plaque B7

TTTAGGGGCGGGGGACGATTCAGCCGAGATCGGGCTTGGGCCCAGAGCATGTTCCAGATCCCAGAGTTTGAGC
CGAGTGAGCAGGAAGACTCCAGCTCTGCAGAGAGGGGCCTGGGCCCCAGCCCCGCAGGGGACGGGCCCTCAGG
CTCCGGCAAGCATCATCGCCAGGCCCCAGGCCTCCTGTGGGACGCCAGTCACCAGCAGGAGCAGCCAACCAGC
AGCAGCCATCATGGAGGCGCTGGGGCTGTGGAGATCCGGAGTCGCCACAGCTCCTACCCCGCGGGGACGGAGG
ACGACGAAGGGATGGGGGAGGAGCCCAGCCCCTTTCGGGGCCGCTCGCGCTCGGCGCCCCCCAACCTCTGGGC
AGCACAGCGCTATGGCCGCGAGCTCCGGAGGATGAGTGACGAGTTTGTGGACTCCTTTAAGAAGGGACTTCCT
CGCCCGAAGAGCGCGGGCACAGCAACGCAGATGCGGCAAAGCTCCAGCTGGACGCGAGTCTTCCAGTCCTGGT
GGGATCGGAACTTGGGCAGGGGAAGCTCCGCCCCCTCCCAGTGACCTTCGCTCCACATCCCGAAACTCCACCC
GTTCCCACTGCCCTGGGCAGCCATCTTGAATATGGGCGGAAGTACTTCCCTCAGGCCTATGCAAAAAGAGGAT
CCGTGCTGTCTCCTTTGGAGGGAGGGCTGACCCAGATTCCCTTCCGGTGCGTGTGAAGCCACGGAAGGCTTGG
TCCCATCGGAAGTTTTGGGTTTTCCGCCCACAGCCGCCGGAAGTGGCTCCGTGGCCCCGCCCTCAGGCTCCGG
GCTTTCCCCCAGGCGCCTGCGCTAAGTCGCGAGCCAGTTTAACCGTTGCGTCACCGGGACCCGAGCCCCCGCG
ATGCCCTGGGGGCCGTGCTCACTACCAAATGTTAATAAAGCCCGCGTCTGTGCC (up primer)

LOCUS NM_032989 986 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens BCL2-associated agonist of cell death (BAD), transcript variant 2, mRNA.**
ACCESSION NM_032989 VERSION NM_032989.2 GI:197116382

```
  1 aactagggcc cggagcccgg ggtgctggag ggaggcggca ggcccgggtc aggggcctcg
 61 agatcgggct tgggcccaga gcatgttcca gatcccagag tttgagccga gtgagcagga
121 agactccagc tctgcagaga ggggcctggg ccccagcccc gcaggggacg ggccctcagg
181 ctccggcaag catcatcgcc aggccccagg cctcctgtgg gacgccagtc accagcagga
241 gcagccaacc agcagcagcc atcatggagg cgctggggct gtggagatcc ggagtcgcca
301 cagctcctac cccgcgggga cggaggacga cgaagggatg ggggaggagc ccagcccctt
361 tcggggccgc tcgcgctcgg cgccccccaa cctctgggca gcacagcgct atggccgcga
421 gctccggagg atgagtgacg agtttgtgga ctcctttaag aagggacttc ctcgcccgaa
481 gagcgcgggc acagcaacgc agatgcggca aagctccagc tggacgcgag tcttccagtc
541 ctggtgggat cggaacttgg gcaggggaag ctccgccccc tcccagtgac cttcgctcca
601 catcccgaaa ctccacccgt tcccactgcc ctgggcagcc atcttgaata tgggcggaag
661 tacttccctc aggcctatgc aaaaagagga tccgtgctgt ctcctttgga gggagggctg
721 acccagattc ccttccggtg cgtgtgaagc cacggaaggc ttggtcccat cggaagtttt
781 gggttttccg cccacagccg ccggaagtgg ctccgtggcc ccgccctcag gctccgggct
841 ttcccccagg cgcctgcgct aagtcgcgag ccaggtttaa ccgttgcgtc accgggaccc
901 gagcccccgc gatgccctgg gggccgtgct cactaccaaa tgttaataaa gcccgcgtct
961 gtgccgccga aaaaaaaaaa aaaaaa //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = complete translation of CDS):

Frame 1: **SRDRAWAQSMFQIPEFEPSEQEDSSSAERGLGPSPAGDGPSGSGKHHRQAPGL LWDASHQQEQPTSSSHHGGAGAVEIRSRHSSYPAGTEDDEGMGEEPSPFRGRS RSAPPNLWAAQRYGRELRRMSDEFVDSFKKGLPRPKSAGTATQMRQSSSWTRV FQSWWDRNLGRGSSAPSQ**
Frame 2: AEIGLGPRACSRSQSLSRVSRKTPALQRGAWAPAPQGTGPQAPASIIARPQAS CGTPVTSRSSQPAAAIMEALGLWRSGVATAPTPRGRRTTKGWGRSPAPFGAAR ARRPPTSGQHSAMAASSGG
Frame 3: PRSGLGPEHVPDPRV

Lung Tumour cDNA Library – Round 9 – Plaque D7

ATCCAGGGGGCCGGGGAC**GAATTCA**AGCGCCACACGAAACTTATACAAAGCTGCACCCTCAGCAGATGATAAT
ATCAAGACACCTGCCGAGCGTCTGCGGGGGCCGCTTCCACCCTCAGCGGATGATAATCTCAAGACACCTTCTG
AGCGTCAGCTCACTCCCCTTCCACCCTCAGCTCCACCCTCAGCAGATGATAATATCAAGACACCTGCCGAGCG
TCTGCGGGGGCCGCTTCCACCCTCAGCCGATGATAATCTCAAGACACCTCCCTTAGCTACTCAGGAGGCTGAG
GCAGAAAAACCACGCAAACCCAAGAGGCAGAGGGCGGCTGAGATGGAACCACCTCCCGAACCCAAGAGGCGGA
GGGTCGGTGACGTGGAACCGTCACGCAAACCCAAGAGGCGGAGGGCCGCTGACGTGGAACCATCATCACCCGA
ACCCAAGAGGCGGAGGGTCGGTGACGTGGAACCGTCACGCAAACCCAAGAGGCGGAGGGCCGCTGACGTGGAA
CCATCATTACCCGAACCCAAGAGGCGGAGGTTGAGCTGAGAAGAGGCCAGTGCACTCAAGCCTGAGCAATAAG
AATAAAACCGAGTAGAACAAAATAAAAAATTCAAAAAACAAAACAAAAGCTTGCGGCCGCACTCGAGTAACTA
GTTAACCCCTTGGGGCCTCTAAACGGGTTTGGAGGGGTTA (up primer)

LOCUS XM_002343430 4062 bp mRNA linear PRI 28-JUL-2011
DEFINITION PREDICTED: **Homo sapiens nuclear pore complex-interacting protein-like 3-like, transcript variant 1 (LOC728888), mRNA.**
ACCESSION XM_002343430 VERSION XM_002343430.3 GI:341915756

CDS pos. 637-4005
```
   1 attatatata gtttatggag agggaagtgc agacataaag aggtgaatta tcttacccag
  61 atcacacagc tgataagtgg tggaggcaga atagaatcta aacagtgtgg ctccggagcc
  //
 541 tcaaacatca ccttgttttt ccactcattc aacacttttt ttttaaatta tctaataggt
 601 tggcactcat catgagcccc tgttctcatt ctgcaa
```
atgg tgaagctctc tattgtcctg
```
 661 
```
accccacagt tcctgtccca tgaccagggc cagctcacca aggagctgca gcagcacgta
```
 721 
```
aagtcagtga catgcccatg cgagtacctg aggaaggtta tcaatactct ggctgaccat
```
 781 
```
catcatcgtg ggactgactt tggtggaagt ccttggttac atattattat tgcgtttccg
```
 //
3361 
```
actccccttc caccctcagc tccaccctca gcagatgata atatcaagac acctgccgag
```
3421 
```
cgtctgcggg ggccgcttcc accctcagcc gatgataatc tcaagacacc ttccgagcgt
```
3481 
```
cagctcactc cccttccacc ctcagctcca ccctcagcag atgataatat caagacacct
```
3541 
```
gccgagcgtc tgcgggggcc gcttccaccc tcagcggatg ataatctcaa gacaccttcc
```
3601 
```
gagcgtcagc tcactccccc tccaccctca gctccaacct cagcagatga taatatcaag
```
3661 
```
acacctgccg agcgtctgcg ggggccgctt ccaccctcag ccgatgataa tctcaagaca
```
3721 
```
cctcccttag ctactcagga ggctgaggca gaaaaaccac gcaaacccaa gaggcagagg
```
3781 
```
gcggctgaga tggaaccacc tcccgaaccc aagaggcgga gggtcggtga cgtggaaccg
```
3841 
```
tcacgcaaac ccaagaggcg gagggccgct gacgtggaac catcatcacc cgaacccaag
```
3901 
```
aggcggaggg tcggtgatgt ggaaccgtca cgcaaaccca agaggcggag ggccgctgac
```
3961 
```
gtggaaccat catcacccga acccaagagg cggaggttga gctgagaaga ggccagtgca
```
4021 ctcaagcctg agcaataaga ataaaaccga gtagaacaaa at //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame):

Frame 1: **SATRNLYKAAPSADDNIKTPAERLRGPLPPSADDNLKTPSERQLTPLPPS**
APPSADDNIKTPAERLRGPLPPSADDNLKTPPLATQEAEAEKPRKPKRQR
AAEMEPPPEPKRRRVGDVEPSRKPKRRRAADVEPSSPEPKRRRVGDVEPS
RKPKRRRAADVEPSLPEPKRRRLS
Frame 2: APHETYTKLHPQQMIISRHLPSVCGGRFHPQRMIISRHLLSVSSLPFHPQ
LHPQQMIISRHLPSVCGGRFHPQPMIISRHLP
Frame 3: RHTKLIQSCTLSR

Lung Tumour cDNA Library – Round 9 – Plaque F7

TTTTCAGGCAGCCTAAGACCCTCCCAGAGCCCAGGGGCTTCACCGCAGACCCCAAGCCATTGAGCACATCACC
CAAAGCAGTGGCCAACATCGCGGACCCCTGTGCCTTGTCACAGATGGGTGCTGGTCCTCAGGCGTTGGGGACA
CTGCTGGGTCGATGGGGTCGGATTCTGCCAGTTTCTGCTCTGCAGCCAAAGATGGTCAGAAGCATTGTCACTT
CAGTAACATCAAGTGCTCAAAGACATGGCAACCGTTCAGTGGTACTTAAGTATTCAAAATATACAACTACAGA
TTCTCTGACAGAAACCAGCACGGGGTCTTCACCTTCATTCACCCCACAGGCGACATGCGAGGGAGAACAGCAT
CTCAGTGGTGATTTCCAAACCAAGCCTTTGTTTTCGGTGTGGGGTTTTGGGGGTTTGCTTTAATGTTTTTGAA
ATTGTAAATGTTGGGCTTTGTATTTTGATGTAAACTGAGCATAATGGCATTTTAGGGCCTGTGACCAAAAATG
AAGCTTGTAACGACCATGGATCTGAATAAACATGTCCTTGCTTCTGAAAAAAAAAAAAAAAAAAAAAAAAAA
AGCTTGGGGCCCCCCCCCAAAAAATAATTAACCCCTTGGGGCCCTCAAAGGGGGTTTGGGGGGGGTTTA (up
primer)

BLAST search strategies:

- No similarity via RNA database: "others, Reference RNA Sequences (refseq_rna)", therefore extend search: outside *Homo sapiens* also results in no hit; more dissimilar sequences brings no hit.
- "others, refseq_genomic" results in 4 human hits, all chromosome 22 genomic contig, featuring protein MICAL-3 isoform 1, Plus/Plus, but gDNA
- compare to Liver Tumour G6

Lung Tumour cDNA Library – Round 9 – Plaque D8

GGAACCGCTAC**GAATTCA**GCCCATAATAATAATACATACATACATAGTACATAGAGAATGTGAGGGGATAGCG
CGCAGTGCTCCCGCTGAATGACTCCACTGCTCGGATTGTCACACCTATGGCATATCACAGATGGCAAAAAGCC
GACAGAGGGAGGCCTTCCACAGTGTCCGTACAGGCCGTGGGAGCACTGGACAAGTCTGCAGTGTCCACAACAC
TAAGAAAAATTAACAGCTTCGAAAAGATCTTAAAGTGATTTTGTGAGCAGGATTCCATTTCCGTTTCTAAGTT
TTTAGATATTACAAAGTACCCATATATATGATAAACACTTAACCCAGATATAAATTTTCTCCTCTTTTAAAAA
ACTCAGTTATGTTTTTGAATAATAATAAAAAATCCACCAAATGCGGGGGAAAAACACCAGTTTAGGAAAAGCC
ACGCTGTGCAACTTTCACAGATAACCACATACGTTGGAGTTGACCCTTCACATTTCTTTTTTTCCAAAATTAG
AGCAAAGAGTCAGCTTAAACAAAAAAAAAAACCCTGAAATTTACAATATGGTGATTAGTTTAAAAAAAAAACA
AGAAGGGCTCTGCAAGGGAGACGCCACAAACCAAGCTTGGAAAGCAAAATCATTTTTGTTTCTCTTTGGCAAC
AACAATAACAAGGAATCTTTTTAGTAAAATGAAGCTAAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCC
TTGGGGCCTCTAAACGGTTTTGGGGGGGGTAAAAAACCAAGTTCCCCGGGGGGGGCCCCACCTTTAGCTCCTT
TTACTAAAAAAATCCCTCGTTTTTGGTGGTTGCCAAAAAGAAACAAAATGATTTTCCTTTCAAGCTGGGTTTG
GGGGTCCCCTCGCAAACCCTTCTGTTTCTTTTTTACACAAACCCAAAATGGAAAATTAGGGGTTTTTTTTTTT
TTTTAACACACCCTGTGCTCTATTTGGGAAAAAAAAAAAGGGGAGGGGCCCCCCCCCGAGGGGGTTTTTTAGA
GAGAGTCCAGGGGGGCTCTCTCACACGGGGTTTCTCCCCCCCGGGGGGGGTTTTTTTTTTATACAACAAAACC
ATATCTTTTTAAAGGAGAGAAAAATATTTTGTGGGCGATGAGGTGTGATATATATATAGGGGGCGCGTGTTTT
ATATTACTACATAA (up primer)

AAAAAAGACTTAATCGAGTGCGGCCGCAAGCTTTAGCTTCATTTTACTAAAAAGATTCCTCGTTATTGTTGTT
GCCAAAGAGAAACAAAAATGATTTTGCTTTCCAAGCTTGGTTTGTGGCGTCTCCCTCGCAGAGCCCTTCTCGT
TTCTTTTTTAAACTAATCACCATATTGTAAATTTCAGGGTTTTTTTTTTTGTTTAACCTGACTCTTTGCTCTA
ATTTTGGAAAAAAAAAAATGTGAAGGGTCAACTCCAACGTATGGGGTTATCTGTGAAAGTTGCACAGCGGGGT
TTTTCCTAAACGGGTGTTTTTCCCCCACATTTGGTGGATTTTTTATTATTATTCAAAAACTTAACTGAGTTTT
TTAAAAGAGGAAAAAATTTATTTCTGGGTTAAGTGTTTATCATATATATGGGTACTTTGTAATATCTAAAAAC
TTAGAAACGGAAATGGAATCCTGCTCACAAAATCACTTTAAAATCTTTTCAAAGCTGTTAATTTTTCTTAGTG
TTGGGGACCCTGCAAACTTGTCCAGTGCTCCCACGGCCTGTACGGACACTGTGGAAGGCCTCCCTCTGTCGGC
TTTTTGCCATCTGTGATATGCCATAGGTGTGACAATCCGAGCAGTGGAGTCATTCATCGGGAGCACTGCACGC
TATCCCCTCACATTCTCTATGAACTATGTATGTATGTATTATTATTATTGGCTTGAATTCGGA
TCCCCGAGCATCACACCTGACTGGAGTACGGACAACCT (dow primer)

LOCUS BC051338 5011 bp mRNA linear PRI 15-JUL-2006
DEFINITION **Homo sapiens endothelial PAS domain protein 1, mRNA (cDNA clone MGC:59860 IMAGE:6305604), complete cds.**
ACCESSION BC051338 VERSION BC051338.1 GI:30410994

Gene in backwards

Lung Tumour cDNA Library – Round 9 – Plaque E8

CGATCCTTTTTCGGCCGAGAC**GAATTCA**GCGTTTGTCCTACAAGGCAAAAGACTCTGAGGAGAAAAAGAAATA
TTGAAAAGGGTGATACATGAAGAACTAGACTCTAATTCTGCTAGAACTCCAACCAGGGGGACCTCCAAGGGGA
AGACACATGCAGTTGATCTAAATTGATGATGATGGTGGCGTGTCCTTTTATAGACAGGGAGTTCCCTGTCACT
GGAGCTTTAGGAGAAATCAGGTGACCGCTTGGCAAGAGTATTGACGTGGAAGAAAGCTCAGGTTCAGATGTGT
TTTCGACTGGAGGTTTATCCAACTGTTTATTTAAAATATGGGTCCAAAACTGCTCTGGCTGTTGGAAATATCG
AGATGAACAACGTAAGGCCTCTGCCCTCTAGAAGATCCCAATCTACCTGATAATGATGGCACTGATGCTTGCT
AGCCAACACTGACTAAGTACTATGTGCTAAGCTCTTTACCTGCATTATTTAATCCTCACAAATTAGCCTCATT
CTGTAGGTCAGGAAACTAAGGTTCAAAGAGGCAGATCATAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACC
CCTTGGGGCCTCTAAACGGGTTTTGAAGGGGTTAAAA (up primer)

CTTACGGTGACTTCTCGAGTGCGGCCGCAAGCTTATGATCTGCCTCTTTGAACCTTAGTTTCCTGACCTACAG
AATGAGGCTAATTTGTGAGGATTAAATAATGCAGGTAAAGAGCTTAGCACATAGTACTTAGTCAGTGTTGGCT
AGCAAGCATCAGTGCCATCATTATCAGGTAGATTGGGATCTTCTAGAGGGCAGAGGCCTTACGTTGTTCATCT
CGATATTTCCAACAGCCAGAGCAGTTTTGGACCCATATTTTAAATAAACAGTTGGATAAACCTCCAGTCGAAA
ACACATCTGAACCTGAGCTTTCTTCCACGTCAATACTCTTGCCAAGCGGTCACCTGATTTCTCCTAAAGCTCC
AGTGACAGGGAACTCCCTGTCTATAAAAGGACACGCCACCATCATCATCAATTTAGATCAACTGCATGTGTCT
TCCCCTTGGAGGTCCCCCTGGTTGGAGTTCTAGCAAAATTAGAGTCTAGTTCTTCATGTATCACCCTTTTCAA
TATTTCTTTTTCTCCTCAGAGTCTTTTGCCTTGTAGGACAAACGCTTGAATTCGGATCCCCGAGCATCACACC
TGACTGGAATACGAACAGCTCC (down primer)

LOCUS CR936804 8487 bp mRNA linear HTC 07-OCT-2008
DEFINITION **Homo sapiens mRNA; cDNA DKFZp781B1548 (from clone DKFZp781B1548).**
ACCESSION CR936804 VERSION CR936804.1 GI:60219715

no CDS
1 tgaattattt ggaaacagtt cacaagtagg catcaaccca ccaactggtg aatgttccat
61 ctttggatat tttgtggata tagagcaatg tgaattcata ccttggtcag atttagttcc
//
6301 actaagatga ttccgagact acccagaatt agtgaacgtt aggaacatgt cagaagaagg
6361 atggaactgg gagagtttgt cctacaaggc aaaagactct gaggagaaaa agaaatattg
6421 aaaagggtga tacatgaaga actagactct aattctgcta gaactccaac caggggacc
6481 tccaagggga agacacatgc agttgatcta aattgatgat gatggtggcg tgtcctttta
6541 tagacaggga gttccctgtc actggagctt taggagaaat caggtgaccg cttggcaaga
6601 gtattgacgt ggaagaaagc tcaggttcag atgtgttttc gactggaggt ttatccaact
6661 gtttatttaa aatatgggtc caaaactgct ctggctgttg gaaatatcga gatgaacaac
6721 gtaaggcctc tgccctctag aagatcccaa tctacctgat aatgatggca ctgatgcttg
6781 ctagccaaca ctgactaagt actacgtgct aagctcttta cctgcattat ttaatcctca
6841 caaattagcc tcattctgta ggtcaggaaa ctaaggttca aagaggcaga tcataagctt
6901 tgaacccacc aagtctagct gactctacac ccatagccta acctcttcct acataaccca
6961 ttagacatga aaaggctacc gagaggcaga tgtgcaccag ggccttttct tatgactcac
//
8401 tggaggttgc agtgggccgg gatcatgccg ctgcactccg gcctgggcaa cagagcgaga
8461 ctccatgtta aaaaaaaaaa aaaaaaa

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined complete CDS):

Frame 1: SAFVLQGKRL
Frame 2: QRLSYKAKDSEEKKKY
Frame 3: SVCPTRQKTLRRKRNIEKGDT

Lung Tumour cDNA Library – Round 9 – Plaque G8

TTTCGGGACCTACAC**GAATTCA**GCATCTGCAAAGCCAGTTTGGCCTCCTGGACCACAAGCACCTAGACCATGA
GGTGGCCAAGCCTGCCCGAAGAAAGAGGCTGCCCGAGATGGCCCAGCCAGTTGACCCGGCTCACAATGTCAGC
CGCCTGCACCGGCTGCCCAGGGATTGCCAGGAGCTGTTCCAGGTTGGGGAGAGGCAGAGTGGACTATTTGAAA
TCCAGCCTCAGGGGTCTCCGCCATTTTTGGTGAACTGCAAGATGACCTCAGATGGAGGCTGGACAGTAATTCA
GAGGCGCCACGATGGCTCAGTGGACTTCAACCGGCCCTGGGAAGCCTACAAGGCGGGGTTTGGGGATCCCCAC
GGCGAGTTCTGGCTGGGTCTGGAGAAGGTGCATAGCATCACGGGGGACCGCAACAGCCGCCTGGCCGTGCAGC
TGCGGGACTGGGATGGCAACGCCGAGTTGCTGCAGTTCTCCGTGCACCTGGGTGGCGAGGACACGGCCTATAG
CCTGCAGCTCACTGCACCCGTGGCCGGCCAGCTGGGCGCCACCACCGTCCCACCCAGCGGCCTCTCCGTACCC
TTCTCCACTTGGGACCAGGATCACGACCTCCGCAGGGACAAGAACTGCGCCAAGAGCCTCTCTGGAGGCTGGT
GGTTTGGCACCTGCAGCCATTCCAACCTCAACGGCCAGTACTTCCGCTCCATCCCACAGCAGCGGCAGAAGCT
TGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTTTTTGAA (up primer)

LOCUS HQ447738 1320 bp DNA linear SYN 21-NOV-2010
DEFINITION **Synthetic construct Homo sapiens clone IMAGE:100071069; CCSB012586 02 angiopoietin-like 4 (ANGPTL4) gene, encodes complete protein.**
ACCESSION HQ447738 VERSION HQ447738.1 GI:312151421

CDS 51-1268

```
   1 ttgttgagca atgctttttt ataatgccaa ctttgtacaa aaaagttggc atgagcggtg
  61 ctccgacggc cggggcagcc ctgatgctct gcgccgccac cgccgtgcta ctgagcgctc
 121 agggcggacc cgtgcagtcc aagtcgccgc gctttgcgtc ctgggacgag atgaatgtcc
 181 tggcgcacgg actcctgcag ctcggccagg ggctgcgcga acacgcggag cgcacccgca
 241 gtcagctgag cgcgctggag cggcgcctga gcgcgtgcgg gtccgcctgt cagggaaccg
 301 aggggtccac cgacctcccg ttagcccctg agagccgggt ggaccctgag gtccttcaca
 361 gcctgcagac acaactcaag gctcagaaca gcaggatcca gcaactcttc cacaaggtgg
 421 cccagcagca gcggcacctg gagaagcagc acctgcgaat tcagcatctg caaagccagt
 481 ttggcctcct ggaccacaag cacctagacc atgaggtggc caagcctgcc cgaagaaaga
 541 ggctgcccga gatggcccag ccagttgacc cggctcacaa tgtcagccgc ctgcaccggc
 601 tgcccaggga ttgccaggag ctgttccagg ttggggagag gcagagtgga ctatttgaaa
 661 tccagcctca ggggtctccg ccatttttgg tgaactgcaa gatgacctca gatggaggct
 721 ggacagtaat tcagaggcgc cacgatggct cagtggactt caaccggccc tgggaagcct
 781 acaaggcggg gtttggggat ccccacggcg agttctggct gggtctggag aaggtgcata
 841 gcatcacggg ggaccgcaac agccgcctgg ccgtgcagct gcgggactgg gatggcaacg
 901 ccgagttgct gcagttctcc gtgcacctgg gtggcgagga cacggcctat agcctgcagc
 961 tcactgcacc cgtggccggc cagctgggcg ccaccaccgt cccacccagc ggcctctccg
1021 taccttctc cacttgggac caggatcacg acctccgcag ggacaagaac tgcgccaaga
1081 gcctctctgg aggctggtgg tttggcacct gcagccattc caacctcaac ggccagtact
1141 tccgctccat cccacagcag cggcagaagc ttaagaaggg aatcttctgg aagacctggc
1201 ggggccgcta ctacccactg caggccacca ccatgttgat ccagcccatg gcagcagagg
1261 cagcctccta cccaactttc ttgtacaaag ttggcattat aagaaagcat tgcttatcaa
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined part of CDS):

Frame 1: SASAKPVWPPGPQAPRP
Frame 2: **QHLQSQFGLLDHKHLDHEVAKPARRKRLPEMAQPVDPAHNVSRLHRLPRDCQELFQVGER QSGLFEIQPQGSPPFLVNCKMTSDGGWTVIQRRHDGSVDFNRPWEAYKAGFGDPHGEFWL GLEKVHSITGDRNSRLAVQLRDWDGNAELLQFSVHLGGEDTAYSLQLTAPVAGQLGATTV PPSGLSVPFSTWDQDHDLRRDKNCAKSLSGGWWFGTCSHSNLNGQYFRSIPQQRQKLAAA LE**
Frame 3: SICKASLASWTTST

Lung Tumour cDNA Library – Round 9 – Plaque D9

ATCCGGCAGTCGGACGGAC**GAATTCA**AGCAGGCGCTGGGGCTGTGGAGATCGGGCTTGGGCCCAGAGCATGTT
CCAGATCCCAGAGTTTGAGCCGAGTGAGCAGGAAGACTCCAGCTCTGCAGAGAGGGGCCTGGGCCCCAGCCCC
GCAGGGGACGGGCCCTCAGGCTCCGGCAAGCATCATCGCCAGGCCCCAGGCCTCCTGTGGGACGCCAGTCACC
AGCAGGAGCAGCCAACCAGCAGCAGCCATCATGGAGGCGCTGGGGCTGTGGAGATCCGGAGTCGCCACAGCTC
CTACCCCGCGGGGACGGAGGACGACGAAGGGATGGGGGAGGAGCCCAGCCCCTTTCGGGGCCGCTCGCGCTCG
GCGCCCCCCAACCTCTGGGCAGCACAGCGCTATGGCCGCGAGCTCCGGAGGATGAGTGACGAGTTTGTGGACT
CCTTTAAGAAGGGACTTCCTCGCCCGAAGAGCGCGGGCACAGCAACGCAGATGCGGCAAAGCTCCAGCTGGAC
GCGAGTCTTCCAGTCCTGGTGGGATCGGAACTTGGGCAGGGGAAGCTCCGCCCCCTCCCAGTGACCTTCGCTC
CACATCCCGAAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTA
AACGGGTTTAAGAGGGTTACTAGAGTGG (up primer)

AAAGATGAACTTACTCGAGTGCGGCCGCAAGCTTTCGGGATGTGGAGCGAAGGTCACTGGGAGGGGGCGGAGC
TTCCCCTGCCCAAGTTCCGATCCCACCAGGACTGGAAGACTCGCGTCCAGCTGGAGCTTTGCCGCATCTGCGT
TGCTGTGCCCGCGCTCTTCGGGCGAGGAAGTCCCTTCTTAAAGGAGTCCACAAACTCGTCACTCATCCTCCGG
AGCTCGCGGCCATAGCGCTGTGCTGCCCAGAGGTTGGGGGGCGCCGAGCGCGAGCGGCCCCGAAAGGGGCTGG
GCTCCTCCCCCATCCCTTCGTCGTCCTCCGTCCCCGCGGGGTAGGAGCTGTGGCGACTCCGGATCTCCACAGC
CCCAGCGCCTCCATGATGGCTGCTGCTGGTTGGCTGCTCCTGCTGGTGACTGGCGTCCCACAGGAGGCCTGGG
GCCTGGCGATGATGCTTGCCGGAGCCTGAGGGCCCGTCCCCTGCGGGGCTGGGGCCCAGGCCCCTCTCTGCAA
AGCTGGAGTCTTCCTGCTCACTCGGCTCAAACTCTGGGATCTGGAACATGCTCTGGGCCCAAGCCCGATCTCC
ACAGCCCCAGCGCCTGCTTGAATTCGGATCCCCCATGATCACACCTGACTGGAATA G (down primer)

LOCUS NM_032989 986 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens BCL2-associated agonist of cell death (BAD), transcript variant 2, mRNA.**
ACCESSION NM_032989 VERSION NM_032989.2 GI:197116382

```
  1 aactagggcc cggagcccgg ggtgctggag ggaggcggca ggcccgggtc aggggcctcg
 61 agatcgggct tgggcccaga gcatgttcca gatcccagag tttgagccga gtgagcagga
121 agactccagc tctgcagaga ggggcctggg ccccagcccc gcaggggacg ggccctcagg
181 ctccggcaag catcatcgcc aggccccagg cctcctgtgg gacgccagtc accagcagga
241 gcagccaacc agcagcagcc atcatggagg cgctggggct gtggagatcc ggagtcgcca
301 cagctcctac cccgcgggga cggaggacga cgaagggatg ggggaggagc ccagcccctt
361 tcggggccgc tcgcgctcgg cgcccccaa cctctgggca gcacagcgct atggccgcga
421 gctccggagg atgagtgacg agtttgtgga ctccttaag aagggacttc ctcgcccgaa
481 gagcgcgggc acagcaacgc agatgcggca aagctccagc tggacgcgag tcttccagtc
541 ctggtgggat cggaacttgg gcaggggaag ctccgccccc tcccagtgac cttcgctcca
601 catcccgaaa ctccacccgt tcccactgcc ctgggcagcc atcttgaata tgggcggaag
661 tacttccctc aggcctatgc aaaaagagga tccgtgctgt ctcctttgga gggagggctg
721 acccagattc ccttccggtg cgtgtgaagc cacggaaggc ttggtcccat cggaagtttt
781 gggttttccg cccacagccg ccggaagtgg ctccgtggcc ccgccctcag gctccgggct
841 ttcccccagg cgcctgcgct aagtcgcgag ccaggtttaa ccgttgcgtc accgggaccc
901 gagcccccgc gatgccctgg gggccgtgct cactaccaaa tgttaataaa gcccgcgtct
961 gtgccgccga aaaaaaaaaa aaaaa //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined complete CDS):

Frame 1: **SSRRWGCGDRAWAQSMFQIPEFEPSEQEDSSSAERGLGPSPAGDGPSGSGKHHRQAPGLL WDASHQQEQPTSSSHHGGAGAVEIRSRHSSYPAGTEDDEGMGEEPSPFRGRSRSAPPNLW AAQRYGRELRRMSDEFVDSFKKGLPRPKSAGTATQMRQSSSWTRVFQSWWDRNLGRGSSA PSQ**
Frame 2: QAGAGAVEIGLGPRACSRSQSLSRVSRKTPALQRGAWAPAPQGTGPQAPASIIARPQASC GTPVTSRSSQPAAAIMEALGLWRSGVATAPTPRGRRTTKGWGRSPAPFGAARARRPPTSG QHSAMAASSGG
Frame 3: KQALGLWRSGLGPEHVPDPRV

Lung Tumour cDNA Library – Round 9 – Plaque G9

CCCGGGTA**GAATTCC**GGGGCTTGGGCCCAGAGCATGTTCCAGATCCCAGAGTTTGAGCCGAGTGAGCAGGAAG
ACTCCAGCTCTGCAGAGAGGGGCCTGGGCCCCAGCCCCGCAGGGGACGGGCCCTCAGGCTCCGGCAAGCATCA
TCGCCAGGCCCCAGGCCTCCTGTGGGACGCCAGTCACCAGCAGGAGCAGCCAACCAGCAGCAGCCATCATGGA
GGCGCTGGGGCTGTGGAGATCCGGAGTCGCCACAGCTCCTACCCCGCGGGGACGGAGGACGACGAAGGGATGG
GGGAGGAGCCCAGCCCCTTTCGGGGCCGCTCGCGCTCGGCGCCCCCCAACCTCTGGGCAGCACAGCGCTATGG
CCGCGAGCTCCGGAGGATGAGTGACGAGTTTGTGGACTCCTTTAAGAAGGGACTTCCTCGCCCGAAGAGCGCG
GGCACAGCAACGCAGATGCGGCAAAGCTCCAGCTGGACGCGAGTCTTCCAGTCCTGGTGGGATCGGAACTTGG
GCAGGGGAAGCTCCGCCCCCTCCCAGTGACCTTCGCTCCACATCCCGAAACTCCACCCGTTCCCACTGCCCTG
GGCAGCCATCTTGAATATGGGCGGAAGTACTTCCCTCAGGCCTATGCAAAAAGAGGATCCGTGCTGTCTCCTT
TGGAGGGAGGGCTGACCCAGATTCCCTTCCGGTGCGTGTGAAGCCACGGAAGGCTTGGTCCCATCGGAAGTTT
TGGGTTTTCCGCCCACAGCCGCCGGAAGTGGCTCCGTGGCCCCGCCCTCAGGCTCCGGGCTTTCCCCCAGGCG
CCTGCGCTAAGTCGCGAGCCACGTTTAACCGTTGCGTCACCGGGACCCGAGCCCCCGCGATGCCCTGGGGGCC
GTGCTCACTACCAAATGTTAATAAAGCCCGCGTCTGTGCAAAAAAAAAAAAAAAAAAAGGCGGCGGGCCGCCG
ACTTCCGAGT (up primer)

LOCUS NM_032989 986 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens BCL2-associated agonist of cell death (BAD), transcript variant 2, mRNA.**
ACCESSION NM_032989 VERSION NM_032989.2 GI:197116382

```
  1 aactagggcc cggagcccgg ggtgctggag ggaggcggca ggcccgggtc aggggcctcg
 61 agatcgggct tgggcccaga gcatgttcca gatcccagag tttgagccga gtgagcagga
121 agactccagc tctgcagaga ggggcctggg ccccagcccc gcaggggacg ggccctcagg
181 ctccggcaag catcatcgcc aggccccagg cctcctgtgg gacgccagtc accagcagga
241 gcagccaacc agcagcagcc atcatggagg cgctggggct gtggagatcc ggagtcgcca
301 cagctcctac cccgcgggga cggaggacga cgaagggatg ggggaggagc ccagcccctt
361 tcggggccgc tcgcgctcgg cgccccccaa cctctgggca gcacagcgct atggccgcga
421 gctccggagg atgagtgacg agtttgtgga ctcctttaag aagggacttc ctcgcccgaa
481 gagcgcgggc acagcaacgc agatgcggca aagctccagc tggacgcgag tcttccagtc
541 ctggtgggat cggaacttgg gcaggggaag ctccgccccc tcccagtgac cttcgctcca
601 catcccgaaa ctccacccgt tcccactgcc ctgggcagcc atcttgaata tgggcggaag
661 tacttccctc aggcctatgc aaaaagagga tccgtgctgt ctcctttgga gggagggctg
721 acccagattc ccttccggtg cgtgtgaagc cacggaaggc ttggtcccat cggaagtttt
781 gggttttccg cccacagccg ccggaagtgg ctccgtggcc ccgccctcag gctccgggct
841 ttcccccagg cgcctgcgct aagtcgcgag ccacgtttaa ccgttgcgtc accgggaccc
901 gagcccccgc gatgccctgg gggccgtgct cactaccaaa tgttaataaa gcccgcgtct
961 gtgccgccga aaaaaaaaaa aaaaaa //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = complete translation of CDS):

Frame 1: **GAWAQSMFQIPEFEPSEQEDSSSAERGLGPSPAGDGPSGSGKHHRQAPGLLWDASHQQ**
EQPTSSSHHGGAGAVEIRSRHSSYPAGTEDDEGMGEEPSPFRGRSRSAPPNLWAAQRY
GRELRRMSDEFVDSFKKGLPRPKSAGTATQMRQSSSWTRVFQSWWDRNLGRGSSAPSQ
Frame 2: GLGPRACSRSQSLSRVSRKTPALQRGAWAPAPQGTGPQAPASIIARPQASCGTPVTSR
SSQPAAAIMEALGLWRSGVATAPTPRGRRTTKGWGRSPAPFGAARARRPPTSGQHSAM
AASSGG
Frame 3: GLGPEHVPDPRV

BLASTp result

8.3.3.4.6 Human breast tumour

Breast Tumour cDNA Library – Round 9 – Plaque C10

CCCTGGGACCTGGGAC**GAATTCA**AGCGAGGCGGCAGGCCCGGGTCAGGGGCCTCGAGATCGGGCTTGGGCCCA
GAGCATGTTCCAGATCCCAGAGTTTGAGCCGAGTGAGCAGGAAGACTCCAGCTCTGCAGAGAGGGGCCTGGGC
CCCAGCCCCGCAGGGGACGGGCCCTCAGGCTCCGGCAAGCATCATTCTCAGGCCCCAGGCCTCCTGTGGGACG
CCAGTCACCAGCAGGAGCAGCCAACCAGCAGCAGCCATCATGGAGGCGC (up primer)

LOCUS NM_032989 986 bp mRNA linear PRI 14-AUG-2011
DEFINITION **Homo sapiens BCL2-associated agonist of cell death (BAD), transcript variant 2, mRNA.**
ACCESSION NM_032989 VERSION NM_032989.2 GI:197116382

```
  1 aactagggcc cggagcccgg ggtgctggag ggaggcggca ggcccgggtc aggggcctcg
 61 agatcgggct tgggcccaga gcatgttcca gatcccagag tttgagccga gtgagcagga
121 agactccagc tctgcagaga ggggcctggg ccccagcccc gcaggggacg ggccctcagg
181 ctccggcaag catcatcgcc aggccccagg cctcctgtgg gacgccagtc accagcagga
241 gcagccaacc agcagcagcc atcatggagg cgctggggct gtggagatcc ggagtcgcca
301 cagctcctac cccgcgggga cggaggacga cgaagggatg ggggaggagc ccagcccctt
361 tcggggccgc tcgcgctcgg cgccccccaa cctctgggca gcacagcgct atggccgcga
421 gctccggagg atgagtgacg agtttgtgga ctcctttaag aagggacttc ctcgcccgaa
481 gagcgcgggc acagcaacgc agatgcggca aagctccagc tggacgcgag tcttccagtc
541 ctggtgggat cggaacttgg gcaggggaag ctccgccccc tcccagtgac cttcgctcca
601 catcccgaaa ctccacccgt tcccactgcc ctgggcagcc atcttgaata tgggcggaag
661 tacttccctc aggcctatgc aaaaagagga tccgtgctgt ctcctttgga gggagggctg
721 acccagattc ccttccggtg cgtgtgaagc cacggaaggc ttggtcccat cggaagtttt
781 gggttttccg cccacagccg ccggaagtgg ctccgtggcc ccgccctcag gctccgggct
841 ttcccccagg cgcctgcgct aagtcgcgag ccaggtttaa ccgttgcgtc accgggaccc
901 gagcccccgc gatgccctgg gggccgtgct cactaccaaa tgttaataaa gcccgcgtct
961 gtgccgccga aaaaaaaaaa aaaaaa //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS):

Frame 1: **SEAAGPGQGPRDRAWAQSMFQIPEFEPSEQEDSSSAERGLGPSPAGDGPSGSG KHHSQAPGLLWDASHQQEQPTSSSHHGG**
Frame 2: ARRQARVRGLEIGLGPRACSRSQSLSRVSRKTPALQRGAWAPAPQGTGPQAPA SIILRPQASCGTPVTSRSSQPAAAIMEA
Frame 3: RGGRPGSGASRSGLGPEHVPDPRV

Breast Tumour cDNA Library – Round 9 – Plaque D10

AGGCATCGGACTGGAC**GAATTCA**GCAGAACAGAATGCTGACCCGCAGGCTGTGACAATGCCTGCCACTGAGAC
CAAAAAGGTCAGCCATGTGGCTGATACGAAGGTCAATACAAAGGCTCAGGAGACTGAGGCTGCACCCTCTCAG
GCCCCAGCAGATGAACCTGAGCCTGAGAGTGCAGCTGCCCAGTCTCAGGAGAATCAGGATACTCGGCCCAAGG
TCAAAGCCAAGAAAGCCCGAAAGGTGAAGCATCTGGATGGGGAAGAGGATGGCAGCAGTGATCAGAGTCAGGC
TTCTGGAACCACAGGTGGCCGAAGGGTCTCAAAGGCTCTAATGGCCTCAATGGCCCGCAGGGCTTCAAGGGGT
CCCATAGCCTTTTGGGCCCGCAGGGCATCAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCC
TCTAAACGGGTTTGGAGGGGTTAGAA (up primer)

TCGATAGACTTCTCGAGTGCGGCCGCAAGCTTGATGCCCTGCGGGCCCAAAAGGCTATGGGACCCCTTGAAGC
CCTGCGGGCCATTGAGGCCATTAGAGCCTTTGAGACCCTTCGGCCACCTGTGGTTCCAGAAGCCTGACTCTGA
TCACTGCTGCCATCCTCTTCCCCATCCAGATGCTTCACCTTTCGGGCTTTCTTGGCTTTGACCTTGGGCCGAG
TATCCTGATTCTCCTGAGACTGGGCAGCTGCACTCTCAGGCTCAGGTTCATCTGCTGGGGCCTGAGAGGGTGC
AGCCTCAGTCTCCTGAGCCTTTGTATTGACCTTCGTATCAGCCACATGGCTGACCTTTTTGGTCTCAGTGGCA
GGCATTGTCACAGCCTGCGGGTCAGCATTCTGTTTCTTGCTTGAATTCGGATCCCCGAGCATCACACCTGACT
GGAATACGACAGCTCCA (down primer)

LOCUS AF126181 1904 bp mRNA linear PRI 04-MAY-1999
DEFINITION **Homo sapiens breast cancer-associated gene 1 protein (BCG1) mRNA, complete cds.**
ACCESSION AF126181 VERSION AF126181.1 GI:4732088

```
CDS 70-1890
   1 tagagaaggc agacgcatcc cgaactcgct ggaggacaag gctcagctct tgccaggcca
  61 aattgagaca tgtctgacac aagcgagagt ggtgcaggtc taactcgctt ccaggctgaa
 121 gcttcagaaa aggacagtag ctcgatgatg cagactctgt tgacagtgac ccagaatgtg
 181 gaggtcccag agacaccgaa ggcctcaaag gcactggagg tctcagagga tgtgaaggtc
 241 tcaaaagcct ctggggtctc aaaggccaca gaggtctcaa agaccccaga ggctcgggag
 301 gcacctgcca cccaggcctc gtctactact cagctgactg atacccaggt tctggcagct
 361 gaaaacaaga gtctagcagc tgacaccaag aaacagaatg ctgacccgca ggctgtgaca
 421 atgcctgcca ctgagaccaa aaaggtcagc catgtggctg atacgaaggt caatacaaag
 481 gctcaggaga ctgaggctgc accctctcag gccccagcag atgaacctga gcctgagagt
 541 gcagctgccc agtctcagga gaatcaggat actcggccca aggtcaaagc caagaaagcc
 601 cgaaaggtga agcatctgga tggggaagag gatggcagca gtgatcagag tcaggcttct
 661 ggaaccacag gtggccgaag ggtctcaaag gctctaatgg cctcaatggc ccgcagggct
 721 tcaaggggtc ccatagcctt ttgggcccgc agggcatcaa ggactcggtt ggctgcttgg
 781 gcccggagag ccttgctctc cctgagatca cctaaagccc gtaggggcaa ggctcgccgt
 841 agagctgcca agctccagtc atcccaagag cctgaagcac caccacctcg ggatgtggcc
 901 cttttgcaag ggagggcaaa tgatttggtg aagtaccttt tggctaaaga ccagacgaag
  //
1741 agtgccagtg ccagtgccag caccagtggt ggcttcagtg ctggtgccag cctgaccgcc
1801 actctcacat ttgggctctt cgctggcctt ggtggagctg gtgccagcac cagtggcagc
1861 tctggtgcct gtggtttctc ctacaagtga gattttagat attg


LOCUS       HSU92544    2064 bp    mRNA    linear   PRI 05-JAN-1999
DEFINITION  Human hepatocellular carcinoma associated protein (JCL-1)
mRNA, complete cds.
ACCESSION   U92544 VERSION     U92544.1  GI:4099968

CDS 67-1887
   1 agaaggcaga cgcatcccga actcgctgga ggacaaggct cagctcttgc caggccaaat
  61 tgagacatgt ctgacacaag cgagagtggt gcaggtctaa ctcgcttcca ggctgaagct
 121 tcagaaaagg acagtagctc gatgatgcag actctgttga cagtgaccca gaatgtggag
 181 gtcccagaga caccgaaggc ctcaaaggca ctggaggtct cagaggatgt gaaggtctca
 241 aaagcctctg gggtctcaaa ggccacagag gtctcaaaga cctcagaggc tcgggaggca
 301 cctgccaccc aggcctcgtc tactactcag ctgactgata cccaggttct ggcagctgaa
 361 aacaagagtc tagcagctga caccaagaaa cagaatgctg acccgcaggc tgtgacaatg
 421 cctgccactg agaccaaaaa ggtcagccat gtggctgata cgaaggtcaa tacaaaggct
 481 caggagactg aggctgcacc ctctcaggcc ccagcagatg aacctgagcc tgagagtgca
 541 gctgcccagt ctcaggagaa tcaggatact cggcccaagg tcaaagccaa gaaagcccga
 601 aaggtgaagc atctggatgg ggaagaggat ggcagcagtg atcagagtca ggcttctgga
 661 accacaggtg gccgaagggt ctcaaaggct ctaatggcct caatggcccg cagggcttca
 721 aggggtccca tagccttttg ggcccgcagg gcatcaagga ctcggttggc tgcttgggcc
 781 cggagagcct tgctctccct gagatcacct aaagcccgta ggggcaaggc tcgccgtaga
 841 gctgccaagc tccagtcatc ccaagagcct gaagcaccac cacctcggga tgtggccctt
 901 ttgcaaggga gggcaaatga tttggtgaag taccttttgg ctaaagacca gacgaagatt
  //
1741 gccagtgcca gtgccagcac cagtggtggc ttcagtgctg gtgccagcct gaccgccact
1801 ctcacatttg ggctcttcgc tggccttggt ggagctggtg ccagcaccag tggcagctct
1861 ggtgcctgtg gtttctccta caagtgagat tttagatatt gttaatcctg ccagtctttc
1921 tcttcaagcc agggtgcatc ctcagaaacc tactcaacac agcactctag gcagccacta
1981 tcaatcaatt gaagttgaca ctctgcatta aatctatttg ccatttcaaa aaaaaaaaaa
2041 aaaaaaaaaa aaaaaaaaaa aaaa


LOCUS       BC091503    2065 bp    mRNA    linear   PRI 15-JUL-2006
DEFINITION  Homo sapiens melanoma antigen family D, 2, mRNA (cDNA clone
MGC:111101 IMAGE:30529545), complete cds.
ACCESSION   BC091503 VERSION     BC091503.1  GI:60551663

CDS 66-1886
   1 gaaggcagac gcatcccgaa ctcgctggag gacaaggctc agctcttgcc aggccaaatt
  61 gagacatgtc tgacacaagc gagagtggtg caggtctaac tcgcttccag gctgaagctt
```

121 **cagaaaagga cagtagctcg atgatgcaga ctctgttgac agtgacccag aatgtggagg**
181 **tcccagagac accgaaggcc tcaaaggcac tggaggtctc agaggatgtg aaggtctcaa**
241 **aagcctctgg ggtctcaaag gccacagagg tctcaaagac cccagaggct cgggaggcac**
301 **ctgccaccca ggcctcgtct actactcagc tgactgatac ccaggttctg gcagctgaaa**
361 **acaagagtct agcagctgac accaagaaac agaatgctga cccgcaggct gtgacaatgc**
421 **ctgccactga gaccaaaaag gtcagccatg tggctgatac gaaggtcaat acaaaggctc**
481 **aggagactga ggctgcaccc tctcaggccc cagcagatga acctgagcct gagagtgcag**
541 **ctgcccagtc tcaggagaat caggatactc ggcccaaggt caaagccaag aaagcccgaa**
601 **aggtgaagca tctggatggg gaagaggatg gcagcagtga tcagagtcag gcttctggaa**
661 **ccacaggtgg ccgaagggtc tcaaaggctc taatggcctc aatggcccgc agggcttcaa**
721 **gaggtcccat agccttttgg gcccgcaggg catcaaggac tcggttggct gcttgggccc**
781 **ggagagcctt gctctccctg agatcaccta aagcccgtag gggcaaggct cgccgtagag**
841 **ctgccaagct ccagtcatcc caagagcctg aagcaccacc acctcgggat gtggcccttt**
901 **tgcaagggag ggcaaatgat ttggtgaagt acctttggc taaagaccag acgaagattc**
 //
1681 **ccaaagccca agagagtggc agtgccagca ctggtgccag taccagtacc aataacagtg**
1741 **ccagtgccag tgccagcacc agtggtggct tcagtgctgg tgccagcctg accgccactc**
1801 **tcacatttgg gctcttcgct ggccttggtg gagctggtgc cagcaccagt ggcagctctg**
1861 **gtgcctgtgg tttctcctac aagtga**gatt ttagatattg ttaatcctgc cagtctttct
1921 cttcaagcca gggtgcatcc tcagaaacct actcaacaca gcactctagg cagccactat
1981 caatcaattg aagttgacac tctgcattaa atctatttgc catttctgaa aaaaaaaaaa
2041 aaaaaaaaaa aaaaaaaaaa aaaaa

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame):

Frame 1: SAE**QNADPQAVTMPATETKKVSHVADTKVNTKAQETEAAPSQAPADEPEPESAAAQSQEN**
QDTRPKVKAKKARKVKHLDGEEDGSSDQSQASGTTGGRRVSKALMASMARRASRGPIAFW
ARRASSLRPHSSN
Frame 2: QQNRMLTRRL
Frame 3: SRTEC

BLASTp results perfect matches with protein products of all three above mentioned gene clones