The problem 1

You are a project team from the bioinformatics group of a large multinational pharmaceutical company. Over the past six months you have been involved in the search for new *drugtargets* for important diseases targeted by the company. Lists of thousands of possible targets identified from micro-array based gene expression studies, and the analysis of EST libraries, has been narrowed down to fifty interesting sequences by extensive bioinformatics analyses for two independent programmes. These sequences have been passed to the molecular biology department for manual assessment by scientists with extensive knowledge of the molecular biology of the diseases and experience in target selection. At a subsequent meeting with the head of molecular biology she presents you with one sequence from each programme.

**> Sequence 1**

MARTTSQLYDAVPIQSTVVLCSCPSPSMVRSQTESSTPPGIPGGSKQGPAMDGTAADPRPGAGSLQHAQPPPEPRKRRPEDFKFGKILGEGTFSTVVLARELASSKEYAIKILEKRHIIKQNKVPYVTRQRDVMTRLDHPFFVKLYFTFQDDEKLYYGLSYAKNGELLKYIRKIGSFNETCTRFYTAEIVTALEYLHGKGIIHRDLKPENILLNEDMHIQITDYGTAKVLSPESKEARADSFVGTAQFVSPELLSEKSACKSSDLWALGCIIYQLVAGLGPFRAGDEYLIFQKIIRLEYDFPEKYFPKARNLVEKLLVLDATKRLGCQEMEGYGPLRAHPFFESVSWENLHEETPPRLTAYLPAMSQDNEDCYGNYDNLLSQFGCMQVSSSSS

**> Sequence 2**

MKSNIIFYFSFFFVYLYYVSCNQSTHSTPVNNEEDQEELYIKNKKLEKLKNIVSGDFVGNYKNNEELLNKKIEELQNSKEKNVHVLINGNSIIDEIEKNEENDDNEENNDDDNTYELDMNDDTFLGQNNDSHFENVDDDAVENEQEDENKEKSESFPLFQNLGLFGKNVLSKVKAQSETDTQSKNEQEISTQGQEVQKPAQGGESTFQKDLDKKLYNLGDVFNHVVDISNKKNKINLDEYGKKYTDFKKEYEDFVLNSKEYDIIKNLIIMFGQEDNKSKNGKTDIVSEAKHMTEIFIKLFKDKEYHEQFKNYIYGVYSYAKQNSHLSEKKIKPEEEYKKFLEYSFNLLNTM

Preliminary analyses show that the first sequence has sequence similarity to the Serine/Threonine protein kinase family, and that the second sequence is Merozoite Surface Protein 7 (MSP7)precursorprotein from the malaria parasite Plasmodium falciparum.

These sequences will go through to high throughput screening with an in-house compound library to identify a potential small molecule *lead compound*. However, the head of molecular biology has recently been to a very interesting presentation at a conference about *structure-based drug design* and heard about successful efforts to design drugs that bind in *active sites* and act as *inhibitors* of key disease related enzymes like the HIV protease and influenza neuraminidase. She knows that the process of experimental determination of protein structure, generally involving the use *of X-ray crystallography* or *NMR* is often time consuming and expensive, but other scientists at the conference told her that protein structure can be predicted. She heard about a program called *PHD* that can predict *protein secondary structure* from the primary sequence with high accuracy, and that a process known as *protein structure prediction by comparative modelling* can be performed on the *SWISSMODEL* WWW site*.*  Also a method called *fold recognition* could sometimes be used in certain circumstances. She would like you to advise whether protein structures can be *predicted* for the above sequences, and if the predictions would be useful in a *rational* approach to the discovery of a lead compound, citing evidence to support your claims.