



## Higher National Unit specification

### General information for centres

**Unit title:** Bioinformatics

**Unit code:** DV9D 35

**Unit purpose:** This is a specialised Unit designed as part of the HND awards in Applied Chemistry and Applied Sciences. Candidates will develop a knowledge of bioinformatics terminology and tools.

On completion of the Unit the candidate should be able to:

- 1 Research historical and current developments in bioinformatics.
- 2 Investigate the ways in which key Internet and computational facilities are used in Bioinformatics.
- 3 Perform DNA sequencing, protein alignment, molecular model viewing and drug discovery.

**Credit points and level:** 1 HN credit at SCQF level 8: (8 SCQF credit points at SCQF level 8\*)

*\*SCQF credit points are used to allocate credit to qualifications in the Scottish Credit and Qualifications Framework (SCQF). Each qualification in the Framework is allocated a number of SCQF credit points at an SCQF level. There are 12 SCQF levels, ranging from Access 1 to Doctorates.*

**Recommended prior knowledge and skills:** Access to this Unit is at the discretion of the centre, however some prior knowledge of Information Technology, Genetics, DNA and protein structure would be an advantage. It is suggested that Units such as DNA Structure and Function (DJ6Y 34), Genetics (DP4P 34) and Information Technology: Applications Software 1 (D75X 34) provide a suitable background for this Unit. Centres may also find that this Unit is suitable to be run in conjunction with a Unit in Medicinal Chemistry.

**Core Skills:** There are opportunities to develop the Core Skill of Communication at Higher in this Unit, although there is no automatic certification of Core Skills or Core Skills components.

**Context for delivery:** If this Unit is delivered as part of a Group Award, it is recommended that it should be taught and assessed within the subject area of the Group Award to which it contributes.

**Assessment:** Outcomes 1 and 2 will be assessed by undertaking a research exercise, while Outcome 3 will be assessed with a practical exercise.

## **Higher National Unit specification: statement of standards**

**Unit title:** Bioinformatics

**Unit code:** DV9D 35

The sections of the Unit stating the Outcomes, knowledge and/or skills, and Evidence Requirements are mandatory.

Outcomes 1 and 2 should be assessed holistically, and the Evidence Requirement for this is detailed in Outcome 2.

Where evidence for Outcomes is assessed on a sample basis, the whole of the content listed in the knowledge and/or skills section must be taught and available for assessment. Candidates should not know in advance the items on which they will be assessed and different items should be sampled on each assessment occasion.

### **Outcome 1**

Research historical and current developments in bioinformatics

#### **Knowledge and/or Skills**

- ◆ Historical developments
- ◆ Current developments

### **Outcome 2**

Investigate the ways in which key Internet and computational facilities are used in Bioinformatics

#### **Knowledge and/or Skills**

- ◆ Telnet
- ◆ Ftp
- ◆ World Wide Web
- ◆ E-mail
- ◆ Usenet
- ◆ On-line databases
- ◆ Roles of network Clients and Servers
- ◆ Supercomputers
- ◆ Cluster computers
- ◆ Grid computing
- ◆ New computational techniques

#### **Evidence Requirements**

The evidence for Outcomes 1 and 2 could be generated by an open-book, unsupervised, assessment in the form of a self researched report. This must report on at least:

Both knowledge and skills in Outcome 1:

## Higher National Unit specification: statement of standards

### Unit title: Bioinformatics

- ◆ one historical development in bioinformatics
- ◆ one current development in bioinformatics

Outcome 2: All uses of Internet and computational facilities to support developments in bioinformatics.

The report should be in an accepted report format and the length of the report should be approximately 1,250 words. Where external sources are used these must be correctly referenced or hyper-linked. Candidates should show that they have fully researched both Outcomes and can evaluate each area.

### Assessment Guidelines

This is a self researched report which should be completed under unsupervised open-book conditions. Centres must take steps to assure authenticity of submission. This could be done by oral questioning of the candidates.

Assessment could be done on an Outcome by Outcome basis, but it is recommended that centres assess Outcomes 1 and 2 in a single assessment event.

### Outcome 3

Perform DNA sequencing, protein alignment, molecular model viewing and Drug discovery

### Knowledge and/or Skills

- ◆ DNA similarity search
- ◆ Protein alignment
- ◆ Molecular structure viewer
- ◆ Drug discovery

### Evidence Requirements

Candidates must complete practical assignments of which they must complete one report detailing the assignment. All knowledge and skills should be covered. Given the nature of the subject, the topic chosen must be compatible with the context of delivery of the rest of the course. This is an open-book assessment and centres should take measures to ensure authenticity of submissions. This could be done in the form of oral questioning.

The supplied sequences, fragments, structures or binding site structures must be designed to allow solutions in a reasonable amount (at the most a couple of minutes) of compute time.

### Assessment Guidelines

For each knowledge and/or skills, the choice of task should be informed by the context of the course which the candidate is undertaking. If the assessment is unsupervised centres must take steps to assure authenticity of submission. This could be done by oral questioning of the candidate.

## Administrative Information

**Unit code:** DV9D 35  
**Unit title:** Bioinformatics  
**Superclass category:** RH  
**Original date of publication:** August 2005  
**Version:** 02 (June 2009)

### History of changes:

Version	Description of change	Date
02	Changes made to standardise assessment guidelines.	03/06/09

**Source:** SQA

© Scottish Qualifications Authority 2005, 2009

This publication may be reproduced in whole or in part for educational purposes provided that no profit is derived from reproduction and that, if reproduced in part, the source is acknowledged.

SQA acknowledges the valuable contribution that Scotland's colleges have made to the development of Higher National qualifications.

Additional copies of this Unit specification can be purchased from the Scottish Qualifications Authority. Please contact the Customer Contact Centre for further details, telephone 0845 279 1000.

## **Higher National Unit specification: support notes**

### **Unit title: Bioinformatics**

This part of the Unit specification is offered as guidance. The support notes are not mandatory.

While the exact time allocated to this Unit is at the discretion of the centre, the notional design length is 40 hours.

### **Guidance on the content and context for this Unit**

This Unit is included as a mandatory Unit in the framework of the HND Chemistry. It may also be suitable to be part of the framework for other HND science awards. It is designed to complement the context of such a framework.

The following notes suggest the content and delivery pace for the Unit.

#### **Outcome 1**

The assessment for this is a self researched report, however it would be beneficial to introduce the nature and range of the subject to candidates, especially a definition of the subject.

Against a design length of 40 hours, it is suggested that a significant portion of time be budgeted for lab work. (In this case of bioinformatics workstation might be the lab.) Candidates are expected to perform another 40 hours of self study as part of the Unit.

Please note that bioinformatics is a large and developing field, partly as a result of the technology churn in the computational engines used to support the work. At any time a paradigm shift may change the specifics of how the computational machinery works. It is not possible to anticipate these changes, but some care has been taken to try and future proof this Unit as far as possible.

Some of this support note makes use of hyper links to various resources on the Internet. These may change over time, but at the time of writing, the search term 'bioinformatics' returns some 23,000,000 'hits'. It seems that the Internet will be a rich source for research on this topic for some time. It is not feasible to give an exhaustive list of resources here, but the links presented here are intended to be used as a start into the subject.

#### **Historical developments in bioinformatics (2 hours)**

The highest profile bioinformatics projects is almost certainly the Human genome project (completed in 2003) A major source of information is the project home page at:

[http://www.ornl.gov/sci/techresources/Human\\_Genome/home.shtml](http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml)

#### **Current developments in bioinformatics (2 hours)**

At the time of writing, (May 2005) the Genome for the Rhesus Monkey had just been released. The home page is at:

<http://www.hgsc.bcm.tmc.edu/projects/rmacaque/>

Other developments in genome research can often be found at:

## **Higher National Unit specification: support notes(cont)**

**Unit title:** Bioinformatics

<http://genome.ucsc.edu/>

A current drug discovery project (with an interesting use of Internet and computing facilities) can be found at the Human Proteome folding project. Their web site is at:

<http://www.grid.org/projects/hpf/about.htm>

### **Outcome 2**

#### **Telnet, Ftp, World Wide Web, e-mail and Usenet (3 hours)**

Again, the Internet is a rich source of information on the technologies used for bioinformatics, many of the programs used are the staple of regular Internet usage, and such technologies as the World Wide Web should need little introduction. Candidates may require more introduction to other technologies. The presence of Graphical front end clients for many of these services may reduce the learning curve associated with the older command line clients.

It would be beneficial to give candidates 'hands on' experience of a selection of these tools, in particular if access to bioinformatics services are available. The way that these tools are used to support bioinformatics should be emphasised, with practical examples when appropriate.

#### **On-line databases (3 hours)**

A number of on-line databases can be reached through the rather useful resource at:

<http://bioinformatics.org/>

A number of on-line databases can be reached through:

[http://www.biomol.net/en/bioinfo\\_dna.php](http://www.biomol.net/en/bioinfo_dna.php)

It is important to emphasise the importance of formatting data to the requirements of the particular on-line database being used.

#### **Roles of network Clients and Servers, Supercomputers, Cluster computers, Grid Computing and new computational techniques (2 hours)**

These topics are fairly standard fare for computing texts, but the distinction between high performance computers (Supercomputers and cluster computers) workstations and personal computers should be discussed, especially in terms of how long some procedures would take on a desktop system compared to a compute engine. It is worth noting the distinction between a workstation and a personal computer is being eroded by the performance increases facilitated by advances in technology.

The distinction between clients and servers should be discussed. This should be both in terms of the physical aspects of servers and more importantly in the roles of the different software products.

## Higher National Unit specification: support notes (cont)

**Unit title:** Bioinformatics

### Outcome 3

This Outcome is largely practical in nature, and the amount of time spent should reflect this. It must be noted that the Unit demands that each of the four practical topics must be taught, so it is suggested that approximately three hours be spent on each of these topics. The rest of the time allowed for the Unit should be spent on the specific details of the software and services to be used.

Aspects of the work would be easier if a dedicated bioinformatics workstation was available for use by a candidate.

Although much of the available software is aligned towards Unix/GNU/Linux platforms, such as that found at:

<http://bioinformatics.org/biobrew/>

Note that some GNU/Linux based bioinformatics workstation distributions are freely downloadable but may require some effort to install and configure.

Other platforms are reported as suitable, such an item is at:

<http://www.nodalpoint.org/node/222>

**DNA similarity search (3 hours)**

**Protein alignment (3 hours)**

**Molecular structure viewer (3 hours)**

**Drug discovery (3 hours)**

**Specific software tools (3 hours)**

There are a number of software tool for performing similarity search and alignment. Many are in the public domain, or other similarly permissive licensing terms. Centres may also have access to commercial software. Some commercial products such as AutoDock may be available to academic institutions if a Software Distribution Agreement is signed by an authorised person.

These software packages include but are not limited to:

- ◆ BLAST
- ◆ AutoDock
- ◆ VAST Search
- ◆ Map Viewer
- ◆ Cn3D
- ◆ ExPASy (Expert Protein Analysis System)
- ◆ HMMER
- ◆ GROMACS
- ◆ EMBOSS.

More information on these and more software tools can be readily obtained on the World Wide Web.

## Higher National Unit specification: support notes (cont)

### Unit title: Bioinformatics

For those who feel more comfortable with paper based resources:

Baxevanis and Ouellette (eds) Bioinformatics — A Practical guide to the Analysis of Genes and Proteins, 2<sup>nd</sup> Edition. Wiley 2001 ISBN 0-471-38390-2

This text covers much of the material required for the Unit, but may not be suitable for undergraduates.

Examples of assessments are detailed below:

- 1 From a supplied sequence the candidate must select appropriate software tools and/or services and perform a DNA similarity search. The query must be appropriately formatted for the tool and/or service selected. The results of the search must form the basis of a report, which demonstrates the candidates ability to select appropriate tools, plan the exercise and evaluate the results obtained.
- 2 From a supplied fragment the candidate must select appropriate software tools and/or services and perform a Protein alignment. The input data must be appropriately formatted for the tool and/or service select. The results of the process must form the basis of a report, which demonstrates the candidates ability to select appropriate tools, plan the exercise and evaluate the results obtained.
- 3 Using an agreed structure the candidate must select appropriate software tools and/or services and produce a viewable model of the selected structure. The input data must be appropriately formatted for the tool and/or service selected. The results of the process must form the basis of a report, which demonstrates the candidates ability to select appropriate tools, plan the exercise and evaluate the results obtained. The report must include either printed output from the viewing tool, or an electronic version of the output suitable for viewing on-line without specialist tools.
- 4 From an agreed target disease and binding site(s), the candidate must select appropriate software tools and/or services and generate a set of discovered drugs that would form the corpus for further screening. The results of the process must form the basis of a report, which demonstrates the candidates ability to select appropriate tools, plan the exercise and evaluate the results obtained. The results of this phase of drug discovery may be voluminous, and may be submitted in electronic format.

### Guidance on the delivery and assessment of this Unit

This Unit will require a mixture of delivery methods. Formal classes will be needed to cover the main theory aspects, while extended laboratory time will be required for Outcome 3. Tutorial support could be used to enhance the learning experience.

#### *Opportunities for developing Core Skills*

There are opportunities to develop the Core Skill of Communication at Higher in this Unit, although there is no automatic certification of Core Skills or Core Skills components.

## **Higher National Unit specification: support notes (cont)**

**Unit title:** Bioinformatics

### **Open learning**

While Outcomes 1 and 2 could be covered via distance learning, it is harder to see how this would be possible for Outcome 3. If a candidate was employed in a laboratory position, then it may be possible for the practical work to be carried out away from the centre. Steps would have to be taken to ensure that the work could be verified as that of the candidate. A blended learning approach may be possible, with candidates only attending college for laboratory work.

### **Disabled candidates and/or those with additional support needs**

The additional support needs of individual candidates should be taken into account when planning learning experiences, selecting assessment instruments, or considering whether any reasonable adjustments may be required. Further advice can be found on our website [www.sqa.org.uk/assessmentarrangements](http://www.sqa.org.uk/assessmentarrangements)

## General information for candidates

### Unit title: Bioinformatics

Bioinformatics is an emerging discipline which combines elements of Biology, (in particular of Cellular Biology), Information Technology (in particular some aspects of Computer Science) and Statistics. At the time of writing there is also an inclusion of aspects of Chemistry. The use of computer technology to support research in Chemistry may in time become more distinct from bioinformatics and become known as chemoinformatics. As the amount of data available to biologists increases, so it becomes more important for the use of computerised analysis tools to increase. With such high profile projects as the Human Genome Project having delivered a full mapping of Human DNA the importance of the use of these tools is becoming more apparent. As bioinformatics only becomes a useful tool when dealing with large amounts of data, this Unit will probably only be delivered when you have been introduced to DNA sequencing, genetics and large protein structures.

This is a specialised Unit, which will be delivered in the second year of the HND Chemistry or HND Applied Sciences courses. It is unlikely to be delivered as a stand-alone Unit, although some colleges may see some merit in having it as an option in other science based or Information Technology HND courses. Bioinformatics is a large subject area, and this Unit can only serve as an introduction to some of the techniques and tools.

On completion of this Unit you should be able to:

- ◆ appreciate the background of bioinformatics
- ◆ understand some important historical developments in bioinformatics
- ◆ investigate current developments in bioinformatics
- ◆ use Internet technologies used to support bioinformatics
- ◆ appreciate computational facilities used to support bioinformatics
- ◆ use software tools and services to perform bioinformatics

### Assessment

There are two assessments for this Unit. The first is a self researched exercised to investigate the history of and current developments in bioinformatics, and the software tools and services used to perform bioinformatics procedures.

The second assessment is to perform a practical assignment. This will involve you selecting and using software tools and/or services to solve a practically based task.