

FROM SCIENTIFIC PROTOCOL TO INFORMATION SYSTEM: the development of an information system supporting bioinformatics analysis

Joan C. Bartlett,

Graduate School of Library and Information Studies, McGill University, joan.bartlett@mcgill.ca

BACKGROUND

- Current need for laboratory scientists to integrate bioinformatics analysis with laboratory science
- Bioinformatics analysis is a complex, multi-step process using multiple resources
- Knowledge about bioinformatics analysis tends to be conveyed by word-of-mouth, limiting its accessibility
- Lack of a systematically developed and documented approach to link bioinformatics analysis to a research task

BIOINFORMATICS ANALYSIS PROTOCOL

- Detailed description of the bioinformatics-based functional analysis of a gene sequence (high level shown in Figure 1)
- Based on semi-structured interviews with 20 bioinformatics experts (7 research groups in Canada and the United States)
- Includes: 12 steps; 3 alternate pathways; detailed description of each step (rationale, input/output data, tools, interpretation, caveats, next steps)
- Involves over 70 individual bioinformatics resources, each with its own unique architecture

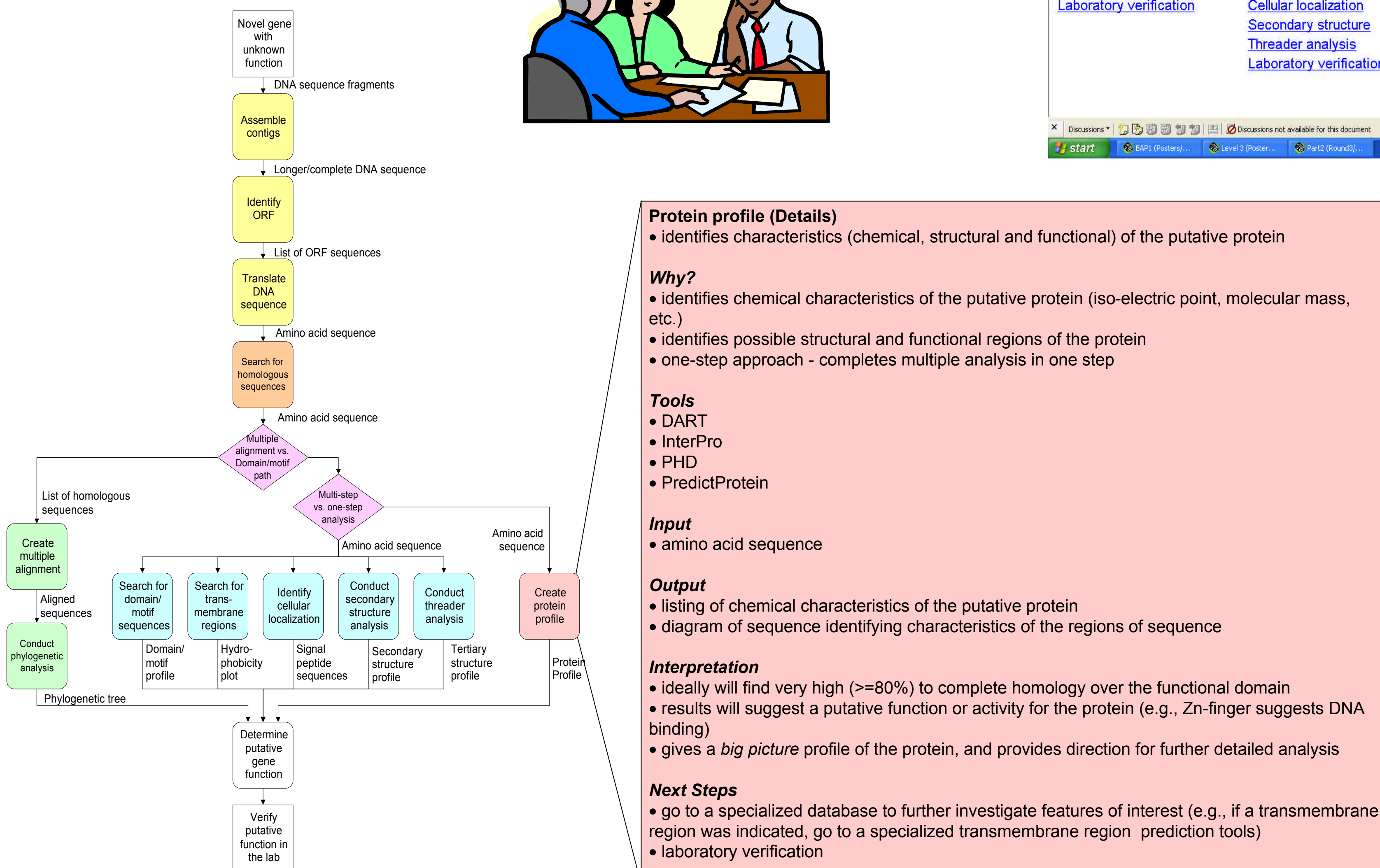


Figure 1. Bioinformatics Analysis Protocol

OBJECTIVE

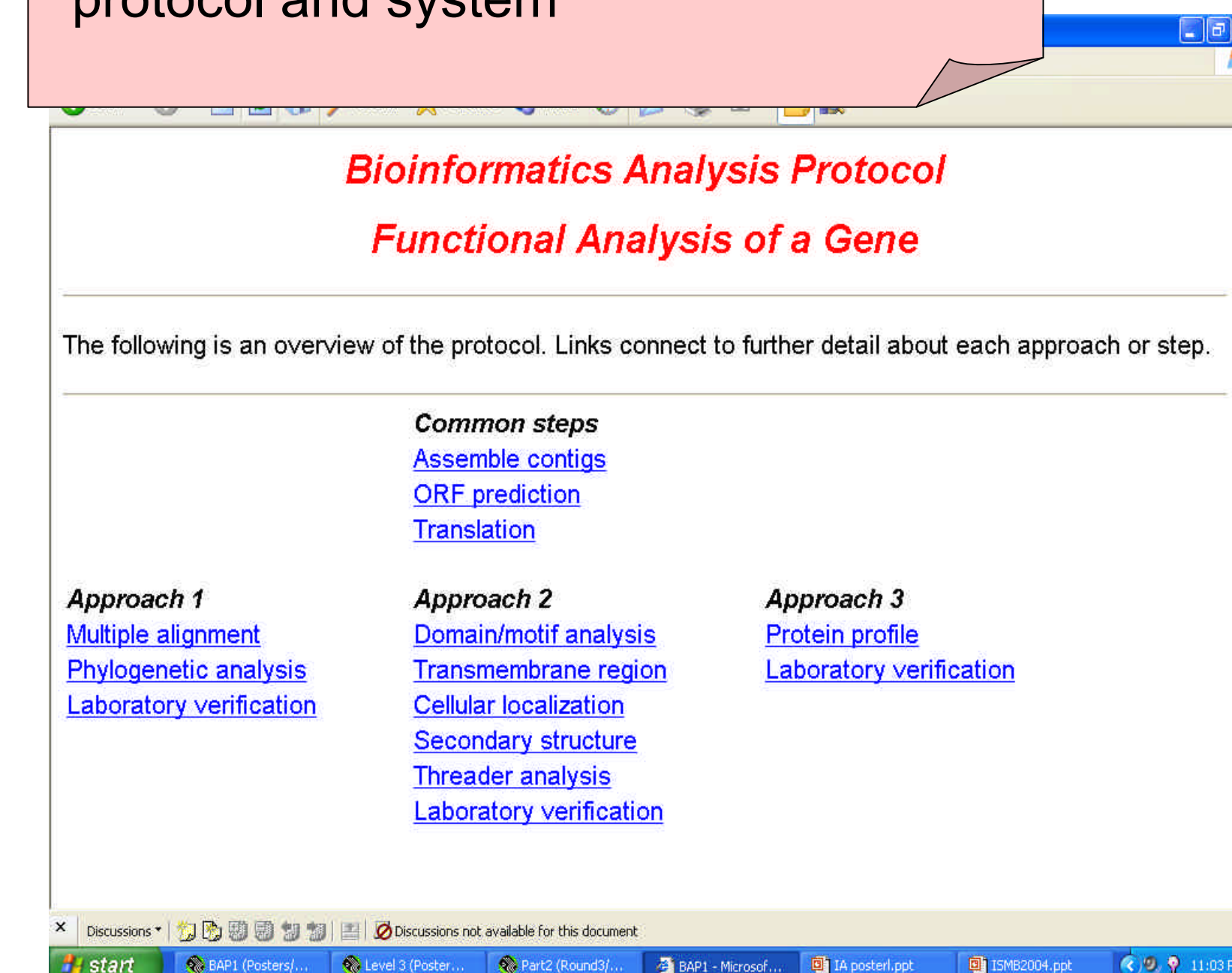
To develop and evaluate a prototype information system to bring the bioinformatics analysis protocol to a laboratory scientist (novice in bioinformatics)

PROTOTYPE INFORMATION SYSTEM

- Designed to guide and direct a laboratory scientist through the functional analysis of a gene sequence
- Follows a hierarchical structure, paralleling the description of the bioinformatics analysis protocol (Figure 2)
- Integrates the use of many different information resources
- Supports the work of scientists unfamiliar with the protocol

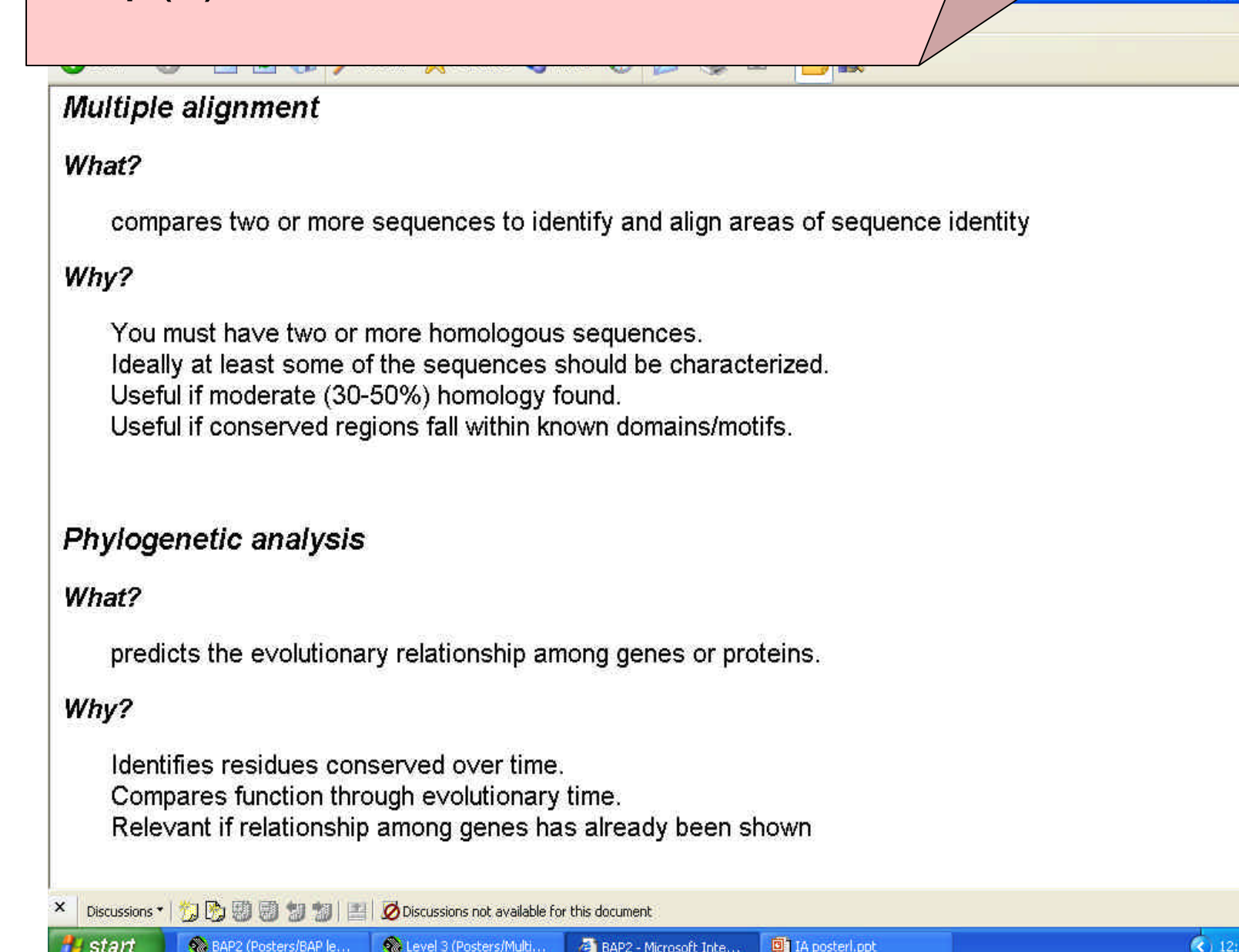
LEVEL 1

- High-level overview of the protocol
- Provides orientation within the protocol and system



LEVEL 2

- Supports navigation between steps
- Provides rationale for each step
- Describes scenario(s) in which each step would be appropriate
- Informs the decision-making process of the scientist in choosing which step(s) to follow



LEVEL 3

- Detailed description of each step
- Includes information on:
 - bioinformatics tool(s) to use (linked to each site)
 - input data (and formatting)
 - interpretation and analysis of results
 - step(s) to follow next
- Provides links to additional information (e.g., further details on a tool, etc.)

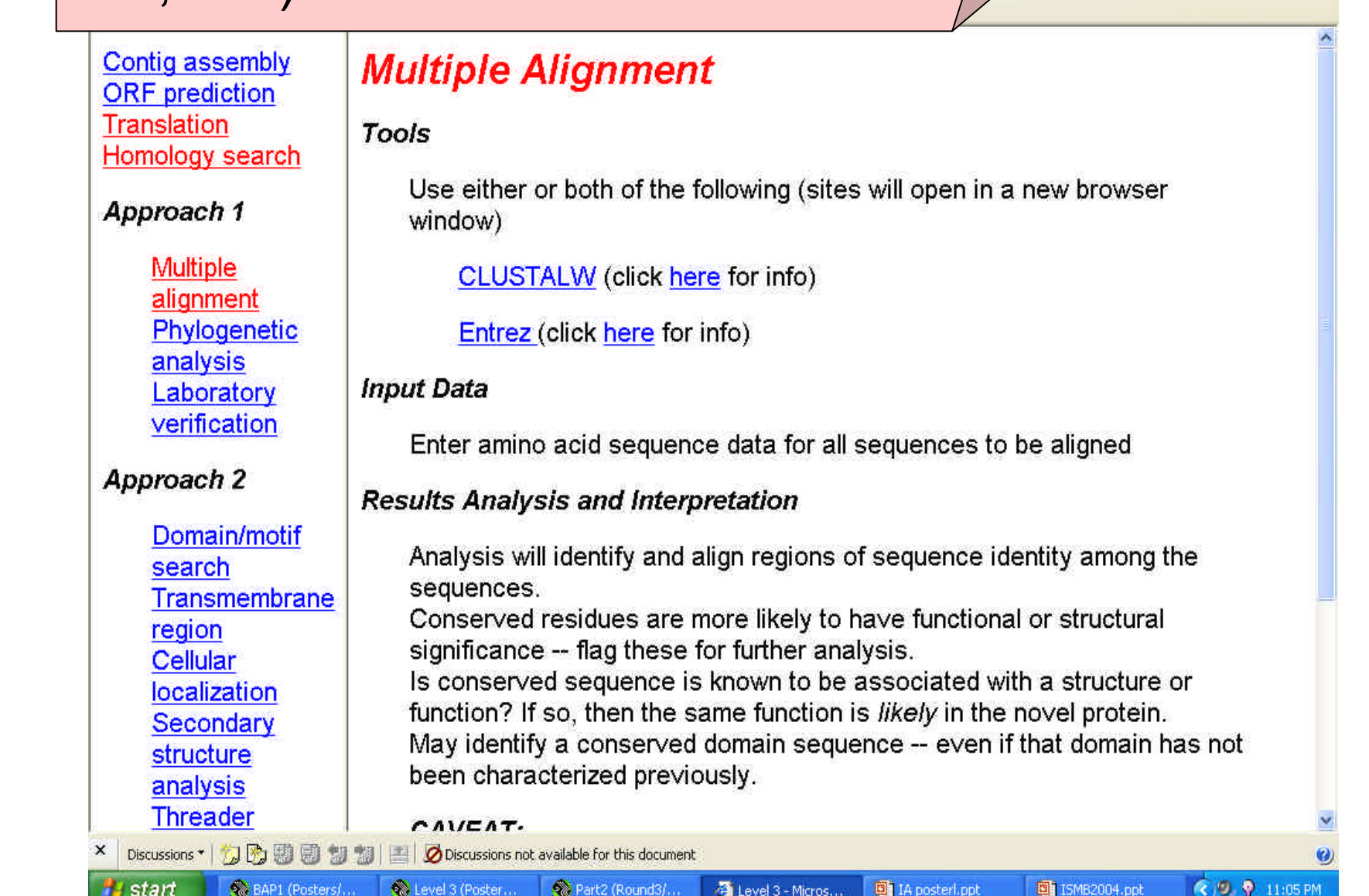


Figure 2. Three Levels of Hierarchy in the Prototype Information System

SYSTEM EVALUATION

- Within subjects experimental design (summer 2005)
- Participants are graduate students in the biological sciences (not bioinformatics specialists)
- Test gene sequence analyzed both with and without use of the information system
- Results of the analyses evaluated and compared, considering factors such as effectiveness, efficiency and usability of the system

ACKNOWLEDGEMENTS

This work is supported by a McGill University New Faculty Start-up Grant

CONCLUSIONS

- Detailed understanding of workflow and dataflow forms the basis for the information system
- Information system integrates and co-ordinates access to and use of over 70 resources
- Provides logical connections among diverse resources
- Follows a task-oriented, user-focused approach to the design of a bioinformatics information system (atypical of most current systems)