Image-Feature Extraction for Protein Crystallization: Integrating Image Analysis and Case-Based Reasoning

I. Jurisica and **P. Rogers**
Ontario Cancer Institute
610 University Avenue
Toronto, ON M5G 2M9
{ij,rogers}@uhnres.utoronto.ca

J. Glasgow and **S. Fortier**
Queen's University
Kingston, ON K7L 3N6
janice@cs.queensu.ca

J. Luft, M. Bianca,
R. Collins, G. DeTitta
Haumptman-Woodward MRI
Buffalo, NY 14203
detitta@hwi.buffalo.edu

Abstract

This paper describes issues related to integrating image analysis techniques into case-based reasoning. Although the approach is generic, a high-throughput protein crystallization problem is used as an example. Our solution to the crystallization problem is to store outcomes of experiments as images, extract important image features, and use them to automatically recognize different crystallization outcomes. Subsequently, we use the outcomes of image classification to perform case-based planning of crystallization experiments for new proteins. Knowledge-discovery techniques are used to extract general principles for crystallization. Such principles are applicable to the adaptation phase of case-based reasoning. The motivation for automated image-feature extraction is twofold: (1) the human interpretation/analysis of image content is subjective, and (2) many problem domains require reasoning with large databases of uninterpreted images. In this paper we present the design and implementation of our integrated system, as well as some preliminary experimental results.

Introduction

This paper describes an application of image analysis techniques to protein crystallization experiment classification. We also describe how this is integrated into our case-based reasoning (CBR) system for protein crystallization experiment design. Image analysis is applicable in many domains that require reasoning about cases (e.g., X-ray interpretation, understanding of NMR images, geographic information systems, satellite image understanding, etc.). Image information plays a crucial role in the domain of molecular biology, where the understanding of the 3D geometry of a molecular structure is often essential to problem solving (Leherte *et al.* 1997). Of particular concern are domains that require image processing without human intervention due to the high throughput (HTP) approaches used for data acquisition.

Case-Based Reasoning With Images A standard technique for human problem solving is to recall past experiences that are in some way similar to the current situation. These experiences, called cases, are then adapted and used to construct a solution for a given problem. CBR systems are computer programs that incorporate such past experiences as a guide to problem solving. CBR may involve adapting old solutions to meet new demands, or using old cases to explain new situations or to critique new solutions.

Cases capture problem-solving processes by storing "important" features of problems and their solutions. Unfortunately, there is no one "right" scheme for representing images as cases; how we choose to represent an image depends on the type of questions we seek to answer. By making particular features of the image explicit, we can provide for efficient pattern matching, retrieval and adaptation in our CBR system. For example, consider the multiple representations of a molecular structure illustrated in Figure 1. If we wish to determine how many atoms of carbon are contained in a molecule, then the formula in Figure 1 a) is sufficient. However, if we need to derive connectivity, angle, distance or shape information, then more complex diagrammatic representations, such as those in Figure 1 b) or c) are more appropriate.

We propose that image information may be stored explicitly (e.g., using a bit map representation) or implicitly (e.g., using shape descriptors that capture some of the shape features of the image). An image may be stored in a way that preserves all relevant visual information, or a simplified model (such as a graph or an array representation) might be the most appropriate form to extract and compare image features. Some form of indexing is required for sizeable image databases where manual indexing is not an option. Following we discuss several issues related to integrating image representations and CBR, including image analysis, automatic image feature extraction, and combining of symbolic and visual information during case retrieval.

CBR with images may involve determining the similarity of images as well as adapting image representations. Psychological studies have provided evidence that suggests the existence of an isomorphism between physical and imagined image transformations (Shepard & Metzler 1971). Similarly, we can propose a set of primitive computational transformation operations that can form the basis for image adaptation. For example, Ohkawa *et al.* (1996) describe a protein classification method using structural transformations, such as deletion, creation, magnification, rotation, movement, exchange or change of kind. In their work, authors compute similarity between proteins on the basis of the cost of individual transformations and their number. Thus, if many

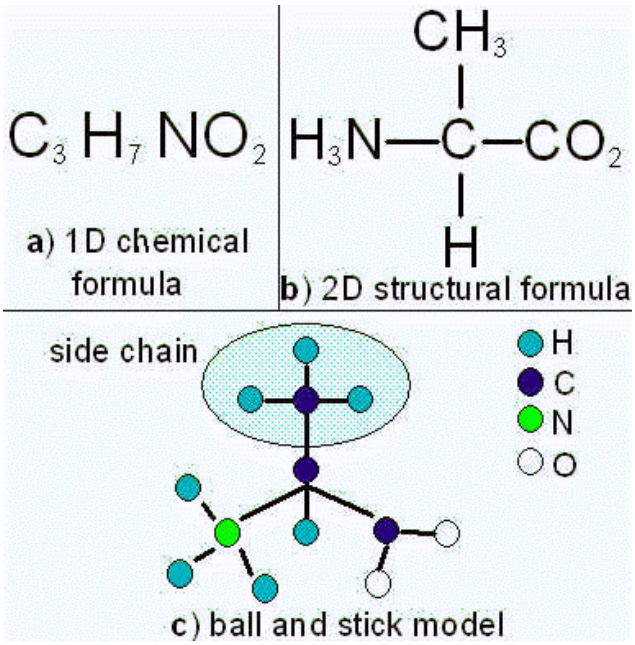

Figure 1: Alternative representations for molecule.

transformations are needed or expensive transformations are required then protein structures are marked dissimilar.

A similar approach has been applied in spatial analogy to the problem of comparing and classifying molecular structures (Conklin *et al.* 1996). The similarity between two images can be measured in terms of the transformations necessary to bring them into equivalence. Considered transformations may include replacing, deleting or moving a part, or rotation of the entire image.

Image Retrieval, Analysis and Use Human interpretation of image content is subjective, and in many domains it may not be feasible due to the complexity of the image or the size of the image database. For such situations, we propose an image-feature extraction system which provides *image segmentation* tools to identify objects within the image, and *image analysis* tools to analyze objects within the image.

Image segmentation has been used to locate objects within an image (Xu, Olman, & Uberbacher 1996) and separate objects during classification (Agam & Dinstein 1997). Popular approaches are based on region-oriented or edge-oriented segmentation. Region-oriented segmentation is based on searching for connected regions with similar gray level values, while edge-oriented segmentation searches for abrupt changes in gray levels that are likely to indicate edges between neighboring objects. An integration of knowledge-based techniques for segmentation is presented in (Tresp *et al.* 1996). Here, a knowledge base is used to determine what objects should be recognized when they have fuzzy boundaries. One can also specify a bias, i.e., domain knowledge about the object.

Image analysis can be used to automatically extract features from images that can subsequently be used for more efficient image retrieval or for decision support. Several approaches have been proposed for the problem of identifying image features: (1) polynomial fitting of flexible curves (McInerney & Terzopoulos 1996; 2000) or planes (Leclerc 1997); (2) attributed relational graphs (ARG) (Petrakis & Faloutsos 1997); (3) similarity invariant coordinate systems (SICS) (Li 1997); and (4) transformational approaches (Basri & Weinshall 1992; Conklin & Glasgow 1992; Ohkawa *et al.* 1996).

Next we introduce a CBR system for protein crystallization. We focus on describing the automated image interpretation and analysis module.

MAX: Protein Crystallization Experience Management System

Proteins are macromolecules which are involved in every biochemical process that maintains life in a living organism. Most disease processes and disease treatments are manifested at the protein level. Through an increased understanding of protein structure we can gain insight into the functions of these important molecules. However, elucidation and understanding of the laws by which proteins adopt their 3D structure is one of the most fundamental challenges in modern molecular biology. Currently, the most powerful method for protein structure determination is single crystal X-ray diffraction. A crystallography experiment begins with a well-formed crystal that ideally diffracts X-rays to high resolution. For proteins this process is often limited by the difficulty of growing crystals suitable for diffraction. This is partially due to the large number of parameters affecting the crystallization outcome (e.g., purity of proteins, intrinsic physico-chemical, biochemical, biophysical and biological parameters) and the unknown correlations between the variation of a parameter and the propensity for a given macromolecule to crystallize.

An ongoing problem in crystal growth is a historically non-systematic approach to knowledge acquisition: *"the history of experiments is not well known, because crystal growers do not monitor parameters"* (Ducruix & Giege 1992, page 14). The Biological Macromolecular Crystallization Database (BMCD) stores data from published crystallization papers, including information about the macromolecule itself, the crystallization methods used, and the crystal data (Gilliland *et al.* 1994). There have been several attempts to analyze the BMCD in order to discover underlying principles of the crystal growth process. These efforts include approaches that use cluster analysis (Farr, Perryman, & Samuzdi 1998; Samuzdi, Fivash, & Rosenberg 1992), inductive learning techniques (Hennessy, Gopalakrishnan, & Buchanan 1994), and statistical analysis (Hennessy *et al.* 2000) to extract knowledge from this existing database of crystallization experiments. Previous studies were limited because negative results are not reported in the database and because many crystallization experiments are not reproducible due to an incomplete method description, missing details, or erroneous data. Consequently, the BMCD is not currently being used in a strongly predictive fashion.

To address these challenges, we are developing MAX, a

case-based reasoning system for the design and evaluation of macromolecular crystal growing experiments. Our objective is that MAX will act as a decision-support system that will incorporate a case base of prior crystal growing experiments to assist an expert crystallographer in the planning of experiments for a novel protein. We extend basic case-based reasoning functionality by providing:

1. image-based processing to extend expressibility of case representation, provide protein similarity measure, and to assure an objective classification of crystallization experiment outcomes with appropriate image-feature extraction;
2. database techniques for case retrieval to support scalability;
3. knowledge-discovery techniques to support domain-knowledge evolution and system optimization.

The repository of crystal growth experiments being developed for our project addresses both of the shortcomings of the BMCD, since it comprises a comprehensive case base of crystallographic experiments that contains both positive and negative experiment outcomes. To build this repository systematically, we combine a HTP crystallization setup and evaluation in the wet lab with computational analysis of the outcomes. We have implemented a scalable, conversational CBR system that uses prior crystal growing experiments to assist an expert crystallographer in the planning of experiments for a novel protein. To support scalability we use an incremental similarity-based retrieval algorithm and the IBM DB2 database system as a back-end storage manager. The information repository contains data and knowledge. Data comprises existing databases (verified information from the Protein Data Bank (PDB) (Bhat *et al.* 2001), the Biological Macromolecule Crystallization Database (BMCD) (Gilliland *et al.* 1994), and GenBank (Database of DNA sequences) (Benson *et al.* 2000)), as well as specialized information about proteins (amino acid sequence, protein properties, etc.), chemicals, and agents.

Knowledge in the system is represented as *cases* – experiments with diverse crystallization outcomes, recorded as a function of time, and *rules* – general principles acquired from crystallographers, or principles derived using knowledge-mining tools. Rules are used to adapt a previous plan to derive a crystallization recipe for a new protein. A mature case base will be used to retrospectively search the cases for interesting and unanticipated relationships. Using data visualization tools and formal knowledge-discovery algorithms for numeric and conceptual cluster analysis, we hope to uncover interesting trends in the outcomes that can be exploited as we face new crystallization challenges.

Essential to building the repository systematically is the automated analysis of experiment outcome. The wet lab uses a robotic HTP crystallization setup with the capacity to prepare and evaluate the results of over sixty thousand (61.4K) crystallization experiments in a work week. This creates a need for an automatic image analysis system. Individual experiments are done in high density microassay plates. Each protein is subjected to 1536 crystallization cocktails, covering a wide range of crystallizing agents. Experiments are evaluated automatically on a computer-controlled XY table with micron positioning accuracy by photographing each well with a 2Mpixels digital camera. The XY stand can accommodate 28 plates of experiments, allowing us to photograph 43,008 experiments in about 9 hours. Each photograph is saved as a JPG image (320 x 320 pixels in RBG). Photographs are taken at several time steps: immediately following setup, one day later, two days later, one week later and two weeks later. Each photograph is analyzed and classified according to the outcome, which can be clear drop, amorphous precipitate, phase separation, microcrystals, crystals, and uncertain outcome. Next we describe the processes of image analysis in more detail.

Image Processing

In addition to the issue of scalability, automated image processing is necessary in the crystallization domain because there is no general solution for quantitatively evaluating reaction outcomes under a microscope. The major weakness of existing scoring methods is the tendency to confuse micro crystalline and amorphous precipitates. To increase objectivity, we have implemented a system to extract image features, and to use them to classify crystallization outcomes. The system runs under Matlab on an IBM RS/6000 SP machine.

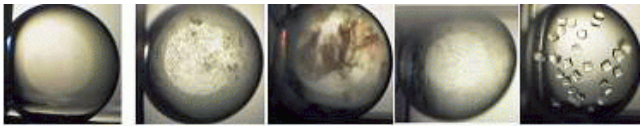

Figure 2: Different crystallization experiment outcomes.

An example of the problem is shown in Figure 2. Image processing is done in four main steps: (1) drop recognition, (2) drop analysis, (3) image-feature extraction, and (4) image classification. Based on image processing, experiment outcomes are classified into appropriate classes to form a precipitation index, which is used to measure similarity between cases during CBR. A precipitation index is a vector with 1536 positions representing crystallization outcomes for one protein and 1536 different crystallizing agents. Its binary representation has 0 for "clear drop" and 1 for "any precipitate" (we also distinguish unknown). A precipitate result can be further broken down into one of amorphous precipitate, phase separation, microcrystals or crystal. During case retrieval, we use both versions of the precipitation index. The binary precipitation index ensures scalability and high recall, while a more detailed precipitation index improves precision in case retrieval.

Drop recognition Drop recognition is performed by locating the well in the image, finding the droplet within the well, and selecting the largest square region contained within the properly illuminated portion of the droplet. The region of interest is then passed on to the drop analysis routine.

The tricky part is identifying the boundaries of the droplet. Our approach generates several feasible outlines of the drop and then uses a weighting mechanism to remove

unlikely candidates. The drop is expected to be of a certain shape, a certain size, and in a certain position (although variations do exist and must be recognized). Figure 3 shows several alternative droplet outlines and Figure 4 presents two examples of processed images, with the oval and square representing the recognized droplets and largest areas of interest respectively. Each image is processed in about 8 seconds, which matches the rate of the HTP capture of crystallization experiment results. Experimental results suggest that the average error rate of the recognition process is $\leq 0.4\%$.

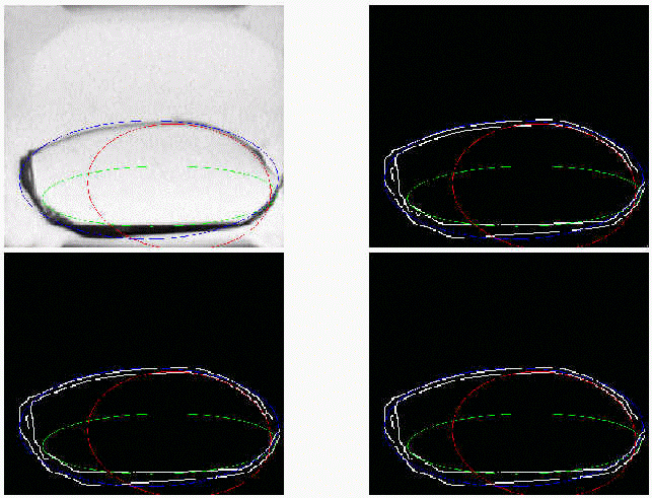

Figure 3: Drop identification process.

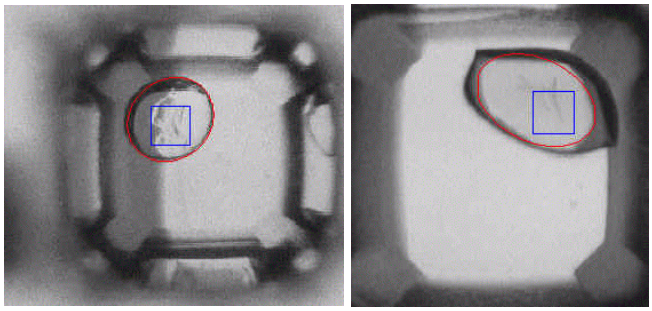

Figure 4: Selecting the area of interest for further image processing.

Drop analysis The goal of analyzing the region of interest is to process the image in a way that enables us to later extract image features that discriminate among different possible crystallization experiment outcomes.

At the moment, we use Fourier transformations to analyze the content of the drop, as seen in Figure 5. In the future, we plan to experiment with alternative techniques, such as wavelet analysis, neural networks, CBR, and combinations of the above approaches.

Image-feature extraction The third step in image processing involves extracting image properties for statistical

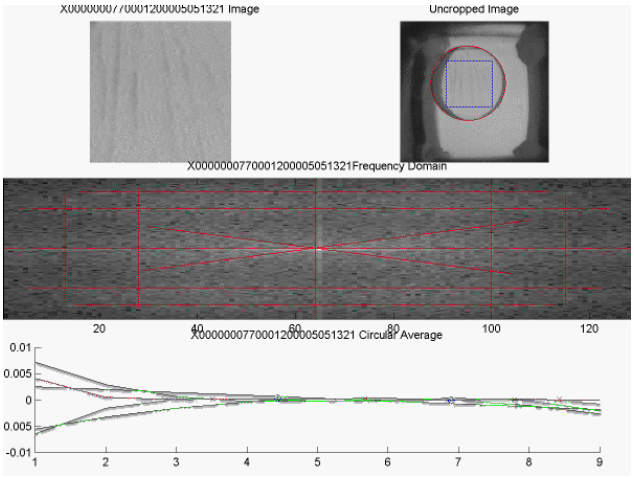

Figure 5: Analysis of the drop content: the recognized drop; the largest square; the spectrum of the Fourier analysis; analysis of the spectrum.

analysis and classification. Currently, we use two types of features: (1) spatial domain features extracted from the 2D intensity map of the image, and (2) frequency domain features extracted from the 2D frequency spectrum of the image.

The spatial domain features correspond to the quadtree decomposition of the image using three different threshold levels. The quadtree decomposition involves splitting the image into four squares and examining the difference between the minimum and maximum values of the pixels in each square. If the difference is greater than a given threshold then the square is further subdivided into four squares, and the quadtree function is called on each square. This process repeats until the minimum and maximum value in each square differs by less than the threshold. We store the number of squares examined.

Other features that are extracted from the spatial domain include the length of edges, the number of bends found in edges, the ratio of the length of edges to their minimum length, and the intensity of edges. We are also experimenting with including the length of edges in a contour plot of the image, and features extracted from the wavelet decomposition of the image.

Frequency domain features are calculated by using a Fourier transformation to find the 2D frequency spectrum of the image, which is then normalized to yield a 2D map of intensity values ranging from 0 to 1. The map is converted to a boolean mask by comparing each value to a threshold constant (currently the value is 0.00007). which is selected empirically so that the shape of the spectrum is well characterized by the boolean mask.

After thresholding, isolated values are filtered out of the mask to simplify the measurements. Measurements are made to parameterize the features of this shape. The height of the horizontal bar of the cross is measured at five different locations chosen to capture the variations in the height of the bar. The width of the vertical bar of the cross is measured

at five different locations, chosen to capture the variations in the width of the bar. The ratio of the height of the horizontal bar to the width of the vertical bar and the ratio of the length of the horizontal bar to the length of the vertical bar are computed. Radial measurements made from the center of the mask to the edge of the cross at varying degrees, along with their variance, are computed and stored as parameters. Finally the number of pixels in the mask is stored in the area parameter.

Features are also extracted from the frequency domain by calculating the circular average. The 2D frequency domain is reduced to a 1D vector by taking the average intensity of all values at different radial distances from the center of the image. The resulting vector is then sampled at three different locations. A fifth-degree polynomial is then fit to the curve and its third derivative and roots are calculated. Currently, about 70 features are extracted to classify experimental outcomes.

Classification of experiment outcomes Extracted features comprise an image description of a case (see Figure 6). Since none of the extracted features is a sole predictor of experiment outcome, the weighted contribution of extracted features is used to automatically classify the outcome of the experiment. We have used CBR to identify the appropriate combination of features and their relative contribution to experiment classification. The example presented in Figure 7 shows the classification of three outcomes. Currently, the accuracy of the experiment outcome classification is 85%.

Classification of crystallization experiment outcomes is used to compute the precipitation index (see Figure 8), which in turn is used to measure similarity between proteins. MAX constructs a solution for the current crystallization problem by using appropriate descriptors from relevant experiments, i.e., both successful and failed crystallization experiments of proteins that have similar precipitation indices. The solution is a recipe for crystallization (i.e, crystal growth method, temperature and pH ranges, concentration of protein, and crystallization agent). Once a novel set of experiments for a protein has been planned, executed and the results recorded, a new case, which reflects this new experience, is added to the case base. Cases with both positive and negative outcomes are equally valuable for future decision-making processes and for the application of machine-learning techniques to the case base.

Figure 6: Part of the case describing the crystallization experiment, which is created using features extracted from the drop.

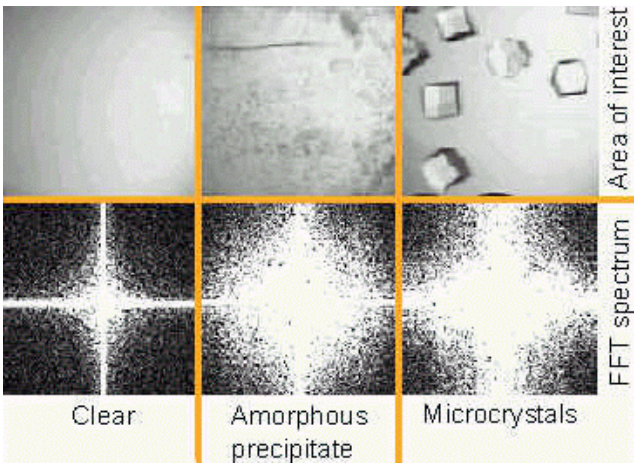

Figure 7: Classification of experiment outcome.

Case-Based Reasoning for Crystallization Experiment Planning

We address the hurdle of protein crystal growth by combining a HTP wet lab work and computational approaches to systematically create and use a comprehensive repository of protein crystallization experiments (Jurisica *et al.* 2000). We apply CBR to plan new crystallization experiments. Our approach is based on a generic system called $\mathcal{TA}$**3** (Jurisica & Glasgow 1997; Jurisica, Glasgow, & Mylopoulos 2000).

Crystallization experiments contain experiential information, such as initial input information about the protein at the beginning of the experiment, the process of carrying out the experiment, the outcome of the experiment, which we rep-

Figure 8: Precipitation index, white represents crystals, light grey (green) is clear drop, black is precipitate, dark grey (red) is unknown.

resent as cases. Thus, we need to address issues of how to represent crystallization experiments flexibly, how to measure similarity among experiments, and how to adapt relevant crystallization recipes to better fit current problem (Jurisica *et al.* 2001).

In general, a *case* represents knowledge of how a specific task was carried out and the outcome for that specific situation. For our domain, an individual case captures the problem-solving process of a crystal growth experiment by representing an episode of such a process: input parameters, results of the initial precipitation experiments, and the final results. Cases are represented as a set of descriptors – attribute-value pairs, organized into attribute categories. Category membership is determined using information about the usefulness of individual attributes and their properties. This information is obtained either from domain knowledge or with help of a knowledge-mining algorithm. Categories bring additional structure to a case representation. This reduces the impact of irrelevant attributes on system competence by selectively using individual categories during matching. A context explicitly defines attributes that are used during similarity assessment and any constraints that may be applicable to attribute values. Thus, context defines how similarity is measured.

As discussed earlier, we use still images to describe individual crystallization experiments. Since our goal is to provide a high-quality information repository cases are linked to external information sources, such as articles describing crystallization methods used, databases of protein information, and chemical properties of agents.

Case retrieval is a primary process for partial pattern matching of an input case to cases in the case base. A similarity function is used to determine which cases are most relevant to the given problem. In MAX case retrieval function is used to locate successful and unsuccessful crystallization experiments that have similar precipitation indexes. The process has two stages. In the first stage, only a binary classification of crystallization outcomes is used (i.e, nothing happened, something happened). In the second stage, a more detailed classification of the result is used to partially order retrieved experiments based on their relevance. Retrieved cases are presented to the user, at which time the user can modify the selection criteria dynamically and thus alter the set of retrieved cases. The retrieval process is interactive and iterative. The retrieval function used in MAX is flexible, effective, and scalable (Jurisica, Glasgow, & Mylopoulos 2000).

The *adaptation* process in CBR manipulates the solution from a set of source cases to solve the target case. MAX constructs a solution for the current crystallization problem by using appropriate descriptors from relevant experiments. The solution is a recipe for crystallization (i.e, crystal growth method, temperature and pH ranges, concentration of protein, and crystallization agent). We propose two functions: *1)* to suggest almost-right solutions to problems, which can be modified automatically (or by the expert user) to suit the new protein situation, and *2)* to warn of potential errors or failures in a proposed experimental plan.

Adaptation is guided by domain knowledge (i.e., adaptation rules, concept hierarchies, or extensive number of examples) that is stored in the MAX information repository or by information provided by the user (in the later scenario, MAX will store the new information for later reuse).

Conclusions

The idea of combining image-based reasoning and CBR is novel, and there are many avenues that need to be explored. Above we have presented just a brief overview of some of the issues involved in integrating these two approaches to reasoning and problem solving. In particular, we have focused on how CBR could be applied in image domains. In some such domains, the combination brings objectivity, in other domains, it is needed to cope with scalability of HTP applications.

We are currently considering the use of CBR in several domains involving image data. In particular, we have considered CBR for the problem of *molecular scene analysis* (Glasgow, Conklin, & Fortier 1993). This work focuses on determining how structural protein data can be organized to permit efficient and rapid retrieval from a case base of molecular scenes. CBR is used to anticipate 3D substructures that might occur within a novel protein image (constructed from an X-ray diffraction experiment).

Medicine is another area with potential for integrating CBR and image-based reasoning. Earlier it was shown that CBR can be successfully applied for prediction and diagnosis in *in vitro* fertilization (Jurisica *et al.* 1998). This system initially worked only with symbolic patient data. Later, more detailed information was collected, including oocyte and embryo images. These images were analyzed by embryologists and the extracted information is used by doctors to potentially provide an explanation of multiple failed implantations. Computer-based image analysis has been used to evaluate morphology and developmental features of oocytes and embryos (including cell number, fragmentation, cellular appearance, zona thickness, etc.) objectively (Jurisica & Glasgow 2000). Although humans can analyze images more flexibly, computer vision techniques help to make the process more objective and precise.

We have implemented MAX using a generic CBR shell called $\mathcal{TA}\mathbf{3}$ in Java 2, with both memory and JDBC drivers. Cases can be stored in a hierarchical manner to support more efficient storage (as one protein may be part of multiple crystallization experiments), improved case retrieval performance, and knowledge discovery through exploiting meaningful structure of case base. A web-based interface and

relational schema to store the information about crystallization experiments has been implemented. We are working on improving its performance and extending its knowledge-discovery capabilities. Currently, knowledge discovery supports only case similarity explanation, and $\mathcal{TA}\mathbf{3}$ optimization by case schema refinement and domain knowledge analysis. Once the repository contains more experiments, we can use knowledge-discovery algorithms to support the extraction of general principles of experimental crystal growing plans. In order to extract principles from the crystallization repository, we apply two steps: searching for patterns and describing their properties. We will use conceptual proximity techniques to organize protein crystallization information into groups that reflect reoccurring patterns. Conceptual clustering methods determine clusters not only by attribute similarity but also by conceptual cohesiveness, as defined by background information. We will use an interactive, context-based, nearest-neighbor clustering that supports explicit background knowledge and works with symbolic attributes. Interactive clustering algorithms prove to be useful especially in evolving domains, such as crystallization. Following group analysis we will apply summarization techniques to describe characteristic properties of the identified clusters. These sets of properties can be used to differentiate individual clusters, to identify associations among the clusters, and to identify relationships between properties and individual items (i.e., associations). In addition, we will explore relationships between the outcomes observed in crystallization experiments and other characteristics of the proteins, such as their sequences, observed biophysical properties, which could also be useful in predicting probable crystallization recipes.

Future work in the crystal growth domain involves the implementation of a distributed storage management system using a robotic tape library attached to the IBM RS/6000 SP, Tivoli storage management system and IBM DB2 EEE database. This is essential to keep up with the increasing volume of image data and to support archiving of important information (we already have over 200GB of compressed images containing crystallization experiment outcomes). By improving the quality of the image capture process, we also hope to improve the current error rate of drop recognition (0.4%) and classification accuracy (85%). Our approach has the potential to significantly reduce the time spent looking for initial conditions. The results of our research may thus eliminate a primary bottleneck in modern structural biology.

Acknowledgments

The computing part of this research is supported in part by the Natural Sciences and Engineering Research Council of Canada, Communications and Information Technology Ontario, and IBM Canada; the wet lab is supported in part by the John R. Oishei Foundation and NASA Grant NAG8-1152. Both labs are supported in part by the NIH grant – Northeastern Structural Genomics Consortium (http://www.nesgc.org).

References

Agam, G., and Dinstein, I. 1997. Geometric separation of partially overlapping nonrigid objects applied to automatic chromosome classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(11):1212–1222.

Basri, R., and Weinshall, D. 1992. Distance metric between 3D models and 2D images for recogntion and classification. Technical Report AIM-1373, MIT, AI Lab.

Benson, D.; Karsch-Mizrachi, I.; Lipman, D.; Ostell, J.; Rapp, B.; and Wheeler, D. 2000. Genbank. *Nucleic Acids Res* 28(1):15–18.

Bhat, T. N.; Bourne, P.; Feng, Z.; Gilliland, G.; Jain, S.; Ravichandran, V.; Schneider, B.; Schneider, K.; Thanki, N.; Weissig, H.; Westbrook, J.; and Berman, H. M. 2001. The PDB data uniformity project. *Nucleic Acids Res* 29(1):214–218.

Conklin, D., and Glasgow, J. 1992. Spatial analogy and subsumption. In Sleeman, and Edwards., eds., *Machine Learning: Proceedings of the Ninth International Conference ML(92)*, 111–116. Morgan Kaufmann.

Conklin, D.; Fortier, S.; Glasgow, J.; and Allen, F. 1996. Conformational analysis from crystallographic data using conceptual clustering. *Acta Crystallographica* B52:535–549.

Ducruix, A., and Giege, R. 1992. *Crystallization of Nucleid Acids and Proteins. A Practical Approach*. New York: Oxford University Press.

Farr, R.; Perryman, A.; and Samuzdi, C. 1998. Re-clustering the database for crystallization of macromolecules. *Journal of Crystal Growth* 183(4):653–668.

Gilliland, G.; Tung, M.; Blakeslee, D.; and Ladner, J. 1994. The biological macromolecule crystallization database, version 3.0: New features, data, and the NASA archive for protein crystal growth data. *Acta Crystallogr* D50:408–413.

Glasgow, J.; Conklin, D.; and Fortier, S. 1993. Case-based reasoning for molecular scene analysis. In *Working Notes of the AAAI Spring Symposium on Case-Based Reasoning and Information Retrieval*, 53–62.

Hennessy, D.; Buchanan, B.; Subramanian, D.; Wilkosz, P. A.; and Rosenberg, J. M. 2000. Statistical methods for the objective design of screening procedures for macromolecular crystallization. *Acta Crystallogr D Biol Crystallogr* 56(Pt 7):817–827.

Hennessy, D.; Gopalakrishnan, V.; and Buchanan, B. G. 1994. Induction of rules for biological macromolecule crystallization. In *ISMB'94*, 179–187.

Jurisica, I., and Glasgow, J. 1997. Improving performance of case-based classification using context-based relevance. *International Journal of Artificial Intelligence Tools. Special Issue of IEEE ITCAI-96 Best Papers* 6(4):511–536.

Jurisica, I., and Glasgow, J. 2000. Extending case-based reasoning by discovering and using image features in IVF. In *ACM Symposium on Applied Computing (SAC 2000)*.

Jurisica, I.; Mylopoulos, J.; Glasgow, J.; Shapiro, H.; and Casper, R. F. 1998. Case-based reasoning in IVF: Prediction and knowledge mining. *Artificial Intelligence in Medicine* 12(1):1–24.

Jurisica, I.; Rogers, P.; Glasgow, J.; Fortier, S.; Luft, J.; Wolfley, J.; Bianca, M.; Weeks, D.; and DeTitta, G. 2000. High throughput macromolecular crystallization: An application of case-based reasoning and data mining. In *Methods in Macromolecular Crystallography*. Kluwer Academic Press. in press.

Jurisica, I.; Rogers, P.; Glasgow, J.; Fortier, S.; Luft, J.; Wolfley, J.; Bianca, M.; Weeks, D.; and DeTitta, G. 2001. Intelligent decision support for protein crystal growth. *IBM Systems Journal* 40(2). To appear.

Jurisica, I.; Glasgow, J.; and Mylopoulos, J. 2000. Incremental iterative retrieval and browsing for efficient conversational CBR systems. *International Journal of Applied Intelligence* 12(3):251–268.

Leclerc, Y. G. 1997. Continuous terrain modeling from image sequences with applications to change detection. In *Workshop on Image Understanding*.

Leherte, L.; Glasgow, J.; Baxter, K.; Steeg, E.; and Fortier, S. 1997. Analysis of three-dimensional protein images. *Journal of Artificial Intelligence Research (JAIR)* 125–159.

Li, S. Z. 1997. Invariant representation, matching and pose estimation of 3D space curves under similarity transformation. *Pattern Recognition* 30(3):447–458.

McInerney, T., and Terzopoulos, D. 1996. Deformable models in medical image analysis: A survey. *Medical Image Analysis* 1(2):91–108.

McInerney, T., and Terzopoulos, D. 2000. T-snakes: topology adaptive snakes. *Med Image Anal* 4(2):73–91.

Ohkawa, T.; Namihira, D.; Komoda, N.; Kidera, A.; and Nakamura, H. 1996. Protein structure classification by structural transformations. In *Proc. of the IEEE International Joint Symposia on Intelligence and Systems*, 23–29.

Petrakis, E. G. M., and Faloutsos, C. 1997. Similarity searching in medical image databases. *IEEE Transactions on Knowledge and Data Engineering* 9(3):435–447.

Samuzdi, C. L.; Fivash, M.; and Rosenberg, J. 1992. Cluster analysis of the biological macromolecule crystallization database. *Journal of Crystal Growth* 123:47–58.

Shepard, R., and Metzler, J. 1971. Mental rotation of three-dimensional objects. *Science* 171:701 – 703.

Tresp, C.; Jager, M.; Moser, M.; Hiltner, J.; and Fathi, M. 1996. A new method for image segmentation based on fuzzy knowledge. In *Proc. of the IEEE International Joint Symposia on Intelligence and Systems*, 227–233.

Xu, Y.; Olman, V.; and Uberbacher, E. C. 1996. A segmentation algorithm for noisy images. In *IEEE Int. Joint Symposium on Intelligence and Systems*, 220–226.