Protein Structure Prediction with Visio-spatial Analogy

Jim Davies (jim@jimdavies.org)
Janice Glasgow (janice@cs.queensu.ca)
Tony Kuo (kuo@cs.queensu.ca)
School of Computing, Queen's University
Kingston, ON, K7L 3N6, Canada

Abstract

We show that visio-spatial representations and reasoning techniques can be used as a similarity metric for analogical protein structure prediction. Our system retrieves pairs of α-helices based on contact map similarity, then transfers and adapts the structure information to an unknown helix pair, showing that similar protein contact maps predict similar 3D protein structure. The success of this method provides support for the notion that changing representations can enable similarity metrics in analogy.

Introduction

It is well known that the right representation greatly facilitates reasoning and there is a growing recognition of the need for intelligent architectures to accomodate a diversity of representations.

The guiding theory of our research is that changing representations allows reasoners to see similarities in one representation type that might be difficult to detect in another. For example, teleological representations of a human face and the front of a car may have very little semantic overlap. In this research we focus on visio-spatial representations. In our example, representing the headlights and eyes as circles, and the grill and mouth as a centrally-located hole allows connections to be drawn between these components.

As people often have visio-spatial experiences when solving problems (Casakin and Goldschmidt, 1999; Farah, 1988; Monaghan and Clement, 1999), an important step in establishing our above theory is to computationally show that visio-spatial representations can be used to solve a variety of problems. In this paper we provide support for this notion in the domain of protein structure prediction, an example of a complex problem-solving domain. We will describe the problem, and then how our system, *Triptych*, uses visio-spatial reasoning on image representations to solve it.

Protein Structure Prediction

A primary goal of molecular biology is to understand the biological processes of macromolecules in terms of their physical properties and chemical structure. Since knowing the structure of macromolecules is crucial to understanding their functions, and all life crucially depends on protein function, an important part of molecular biology is understanding the three-dimensional (3D) structure of proteins.

Proteins are composed of one or more chains of amino acid residues. The description of which residues appear and in what order is the protein's "primary structure". According to the laws of chemistry, the chains twist, fold, and bond at different points, forming a complex 3D shape. Subchains form regular "secondary structures", the two main types being α-helices and β-strands. The three-dimensional structure of a chain is its "tertiary structure", and the overall protein shape (which may involve several chains) is known as its "quaternary structure". A major unsolved problem for the biological sciences is to be able to reliably predict the quaternary structure from the primary. This, at the highest level, is our problem domain.

Approaches to protein structure prediction vary from those that apply physical principles to those that consider known amino acid sequences and previously determined protein structures. Many of the latter use what is known as "homology" as a similarity metric. In this context homology is the similarity of two amino acid sequences. Our work also falls in the latter category, but rather than using primary structure directly, we compare contact maps.

Contact maps A *distance map*, D, for a protein with n amino acid residues is an $n \times n$, symmetric array where entry $D(a_i, a_j)$ is the distance between residue a_i and residue a_j, generally calculated at the coordinates of the C_α (carbon-alpha) atoms for the residues. Given a distance map D, we compute a *contact map* C for the protein as a symmetric, $n \times n$ array such that:

$$C(a_i, a_j) = \begin{cases} 1, & \text{if } D(a_i, a_j) < t; \\ 0, & \text{otherwise.} \end{cases}$$

where t is a given threshold value (in our work this theshold is $10\mathring{A}$). There exists a contact between residues a_i and a_j if and only if they are within a given distance t of one another in the protein structure.

Researchers have considered various approaches for the process of predicting contact maps for a protein from its primary sequence and structural features; these are primarily based on neural network-based methods (Fariselli et al., 2001). While results from this work is encouraging, it still results in maps that contain a large degree of noise. Thus we carry out our initial experiments on idealized maps generated from the Protein Data Bank (PDB) (Berman et al., 2000). Future work

will include prediction of structure from predicted contact maps.

A contact map is a translational and rotational invariant, visio-spatial representation that captures some of the protein's relevant structural information. Our general hypothesis is that visual processing on contact maps enables effective retrieval of similar structures, even if homology sequence is ignored. Contact maps provide a "fingerprint" that can be used to efficiently compare proteins to find ones with similar substructures. We will refine this hypothesis when we describe our implementation.

Analogy Applied to Protein Structure Prediction

Rather than working with whole proteins, we are working with pairs of α-helices. At the highest level, each time Triptych runs it takes as input: 1) the contact map for the unknown (target) helix pair, and 2) a memory of known helix pair structures and contact maps. The final output consists of a location in space (*x, y, z* coordinates) of each amino acid residue in the target helix pair.

Analogical problem solving is founded on the premise that similar problems have similar solutions. Experiences are retrieved, mapped, and reused during problem solving. Aaronson et al. (Aaronson et al., 1993) suggest that analogical reasoning is particularly applicable to the biological domain, partly because biological systems are often homologous (rooted in evolution). As well, biologists often use analogy, where experiments are designed and performed based on the similarity between features of a new system and those of known systems. Analogical and case-based reasoning has previously been applied to a number of problems in molecular biology; an overview of these systems can be found in (Jurisica and Glasgow, 2004).

Our system retrieves and adapts protein data from the PDB in order to construct potential 3D structural models for our target helix pair. These models are evaluated in terms of domain knowledge and the "best" structures will ultimately be used as building blocks at the next level of model building.

We retrieve similar α-helix pair contact maps and adapt the known structures to predict alignments for the unknown structures.

To predict the alignment of helices in 3D space, we consider helix pair contact maps, C_{s_m,s_n}, corresponding to pairs of helices (s_m, s_n) such that there are greater than four contacts..[1] This map is the subarray of C such that the the rows of C_{s_m,s_n} correspond to the amino acid residues in secondary structure s_m and the columns correspond to the residues in secondary structure s_n. These maps need only be defined for contacts along and below the diagonal of the helix pair contact map, as the map for pair (s_m, s_n) is equivalent to that for (s_n, s_m). Note, that unlike the protein contact map the contact maps for pairs of helices are not generally symmetric. The images in Figure 1 illustrate a contact maps for pairs of α-helices.

The *retrieve* task returns a list of retrieved helix pairs, ordered according to similarity. The similarity metric is a visual similarity between source and target contact maps. The *adapt* module transfers structure information from the top retrievals (called the "sources") and modifies the information according the the specifics of the target.

Implemented Modules: Retrieval and Adaptation

Our focus is on predicting the alignment, or relative location, in 3D space of α-helix pairs given the contacts between their residues.

Retrieval Module For each query map C_{s_m,s_n} we retrieve helix pairs with contact maps most similar to C_{s_m,s_n}.

A similarity measure for comparing the query contact map with maps generated from structures in the PDB was derived using techniques from machine vision, where we consider the black regions to be the image within the array. We were less concerned about the dimensions of the map, than what it looked like in terms of shape and location of black regions (regions which contain contacts). For example, Figure 1 illustrates three different maps for pairs of helices, where maps (a) and (c) are considered similar to one another, and (b) is different from the other two.

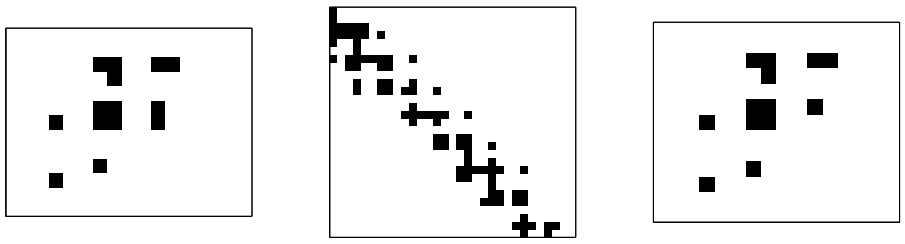
First we blur the images using Gaussian smoothing (Gonzalez and Woods, 1992). This is often done to remove unwanted details and noise. Contacts are treated as black points, and points surrounding them are turned some shade of gray depending on their distance from the nearest contacts. The grayscale tone is determined by a Gaussian distribution where the contacts are the means. The maps are then morphed using a technique called *closing,* which removes low-valued points but keeps the rest of the image intact (Gonzalez and Woods, 1992).

The retrieval of similar contact maps involves a two-tiered approach. Given a query contact map, the first tier uses three general content descriptors to cull the dataset of dissimilar contact maps: quadtrees, color and edge distributions, and gray-level co-occurrence matrices.

Quadtrees have been successfully applied to image compression, comparison, and classification. The quadtree (Sullivan and Baker, 1994) is a hierarchical data structure used to represent images. For an image, a two-dimensional region is recursively decomposed into quadrants where each quadrant is a node in the quadtree.

Color distribution (Smith and Chang, 1994) is a common feature used in image retrieval. Pixel color values are put into a histogram form: colors are discretized and counted and placed in bins. Global histogram representation has the drawback of loss of location, shape, and texture information. As a result images retrieved based on similar color distributions may not be semantically

[1] If there are fewer than five contacts between two secondary structures it is difficult to determine their orientation from their contacts.



(a) (b) (c)

Figure 1: Illustration of similar, (a) and (c), contact maps and a map (b) that is dissimilar to the other two. (b) shows the sub-contact $C_{Helix-6,Helix-8}$ map for a pair of helices in protein Bacterioferritin. Since the diagonal band shows contacts that extend from the beginning of one and end of another, to the end of one and beginning of another, we can discern that the helices are oriented anti-parallel to one another.

related.

Edge detection (Won et al., 2002), and the features that can be extracted from it, is commonly used as a content descriptor of images. In this work we use the Canny edge detection method (Canny, 1986). The Gaussian smoothing was necessary for this step to work, as it uses gradients and cannot be applied to binary images. Our measure of similarity based on edge detection involves comparing histograms showing the frequency of edges with angles of 0^o, 45^o, 90^o, and 135^o.

A statistical method that considers the spatial relationship of pixels, the gray-level co-occurrence matrix (GLCM) (Haralick et al., 1973) is a texture analysis method from which various statistical features can be extracted. Each entry *(i, j)* in the GLCM corresponds to the number of occurrences of the pair of gray levels i and j which are a distance d apart in the original image. For example, if d is 1, then GLCM entry (1, 2) will contain the number 4 if there are four instances of gray value 1 adjacent to gray value 2 in the original image. In analysis, the GLCM are normalized so the histogram or features extracted can be compared.

A committee of these general content descriptors is used in the first tier of retrieval. Quadtrees vectors were generated from the binary, smoothed, and morphed contact maps. The color and edge distributions and gray level co-occurrence matrices were obtained from the smoothed contact maps. The committee results in a set of contact maps which are present in the retrievals of two or more general content descriptors. We determined empirically that 100 retrievals for each descriptor is sufficient. The results of the committee are then used in the second tier of retrieval.

For the second tier, the Jaccard's distance (Jaccard, 1908) was calculated between each contact map from the first tier and the query map. Because the maps vary in size, a sliding window approach was used to determine the best matching regions between the query and the contact maps from the first tier. The best mapping regions also provide registration of residues for evaluation using RMSD (Root-Mean-Squared Distance), the standard measure of distance for protein structures. The best 25 retrievals were then selected from the 100 as the final set of contact maps to be returned.

Adaptation Module The retrieval process returns, for each query contact map, potential helix pairs from the PDB, ranked in order of estimated similarity. For each query map, the adaptation phase transfers the structure information from the highest-ranking structures to the input helix pair.

Transferring locations requires a mapping function – that is, a set of alignments that determine which residues in the target structure map to which residues in the retrieved source structure. This is achieved by first aligning the contact maps so that the mean cell location of contacting amino acid residues in the retrieved structure aligns with the mean cell location of contacting residues in the target. Then all amino acid residues in the target structure that have corresponding residues in the source structure are given the coordinate informa-

tion from these residues. Since the registration may not feature a great overlap of the maps, usually there remain some target residues with no coordinates (i.e., no corresponding residue in the known structure to transfer over). Since α-helices tend to have a consistent structure, the missing coordinates are filled in using general domain knowledge. Specifically, each turn of an α-helix is estimated at 5.4 Å along the helix axis and each turn at 5 Å across. Using this information and the helix axis, calculated from the filled-in locations, our system is able to infer these unmatched residue locations. This is textbook Biochemistry knowledge. Figure 2 illustrates the portions of the helices that are determined through our mapping function and those constructed from domain knowledge (grown area).

Figure 2: In this figure the lower helix is the target and the upper is the source. The dotted gray circle represents the mapping area. The locations of the target amino acid residues for which there are no cooresponding source residues are inferred based on the known geometry for helices. These "grown" areas are represented with the dotted black line.

Given this implementation and our overall hypothesis, our refined hypothesis is that analogy using contact map similarity can effectively generate accurate protein substructure predictions. We applied the retrieval and adaptation components of Triptych to a set of 61 proteins, mostly all α chains, retrieved from the PDB.[2]

Results

For each protein, we computed the distance map, contact map and secondary structure contact map. From the contact maps, we were able to derive 422 maps that described contacts for pairs of helices.

The results of the retrieval process for 422 unique test queries are shown in Table 1. N is the number of helix pairs retrieved; *Mean* describes the average RMSD for the queries and *Std* is the average standard deviation. *Mean Best* and *Rank* describe the average best RMSD and its median rank within the final set of contact maps. The results suggest the following: 1) as N, the number of retrieved helix pairs, decreases the average RMSD of the final set of contact maps improves, 2) the *Mean Best* represents the best structure match and worsens as N decreases, and 3) as N increases from 25 to 50 to 100, the *Mean Best* does not change significantly.

N	$Mean$	Std	$MeanBest$	$Rank$
100	1.8604	0.8035	0.5259	7.5
50	1.6498	0.6447	0.5303	7
25	1.3944	0.5077	0.5506	5
10	1.1919	0.4166	0.6034	3

Table 1: The retrieval results of the committee on 422 unique queries when the top N out of 100 are returned as the final set of contact maps.

n	RMSD
1	3.6668
5	2.2667
10	1.8814
25	1.5286
50	1.3921
100	1.3011
200	1.2507
422(all)	1.2426

Table 2: Experimental results when considering the adaptation of the top N results. RMSD denotes the mean of the best scores for each of the 422 input helix pairs for the top N retrievals.

Further examination of the 100 retrievals using the committee determined that 65.40% of the 422 queries have its best RMSD fall within the top 10 retrievals, 83.18% within the top 25 and 96.45% within the top 50. Thus, a final set of contact maps consisting of the top 25 retrievals from a set of 100 seems to be the best balance between a low average RMSD over all the retrievals and a low RMSD for the average best retrieval. This ensures all the retrievals are similar to the query and contains the best match in $\sim 83\%$ of the helix pairs.

Using the results of the retrievals module, we evaluated the adaptation method by comparing the *predicted* locations of the residues to the *actual* locations, as given in the Protein Data Bank (PDB) in terms of RSMD. The results when considering the top N retrievals, for N = 1, 5, 10 25, 50, 100, 200, and 422 are presented Table 2. These results suggest that we converge to a good solution when considering about the top 50 solutions.

The RMSDs presented are acceptably accurate in the biochemistry literature. See (Wang et al., 2005) for an empirical study with an RMSD of 1.6.

Note that the retrieval scores for the *Mean Best* (in terms of RMSD distance between the correct and predicted structures) are less than the adaptation scores (which reported the distance between the retrieved struc-

[2]The proteins were 1a0aA, 1a1z_, 1a28A, 1acp_, 1afrA, 1aj8A, 1akhA, 1akhB, 1am9A, 1aoiA, 1aoiB, 1arv_, 1auiB, 1auwA, 1bbhA, 1bcfA, 1bgp_, 1bh9A, 1bh9B, 1bu7A, 1bvb_, 1c52_, 1cc5_, 1cem_, 1cktA, 1cll_, 1cpq_, 1csh_, 1cy5A, 1d9cA, 1dceB, 1dpsA, 1ea1A, 1eerA, 1eteA, 1fce_, 1fgjA, 1ft1B, 1furA, 1gakA,
1hcrA, 1hnr_, 1hryA, 1huuA, 1hyp_, 1kx2A, 1lbd_, 1lfb_, 1lis_, 1lmb3,
1mhyD, 1neq_, 1pbwA, 1pru_, 1rzl_, 1tc3C, 1tx4A, 1uxc_, 2af8_, 2hddA, and 2ilk_.

tures and the correct structure). The reason for this is that the retrieval scores are based on the RMSD of only the regions of the helices in contact with each other. The adaptation method extends the helices beyond the regions of contact based on biochemical knowledge, affording more opportunity for error.

Related Work

Previous methods for the recovery of 3D structure from distance contact maps are mostly based on distance geometry and stochastic optimization techniques, though none look specifically at prediction of helix pair structures.

The issue of visual knowledge in analogy and case-based reasoning has attracted the attention of researchers in several areas. Below we relate our work to some analogical problem solving systems that use visio-spatial knowledge.

Previous visual analogy work in molecular biology domains include visualizing crystallographic data at different resolutions (Glasgow et al., 1993), in drug design (Biname et al., 2004), and in in-vitro fertilization (Jurisica and Glasgow, 2000). Perner has applied visual analogy to image segmentation of CT images (Perner, 1999), HEp-2 cell images (Perner, 1998), and the identification of fungi (Perner et al., 2003).

Analogy with spatial reasoning has been applied to non-bioinformatics domains as well. FABEL (Gebhardt et al., 1997) is an example of a system that adapts diagrammatic cases in the domain of architectural design. REBUILDER (Gomes et al., 2003) is an analogical reasoner that does retrieval, mapping, and transfer of software design class diagrams. FAMING (Faltings and Sun, 1996) makes analogies with physical mechanism parts.

Visual analogy has been used for cognitive modeling as well. DIVA (Croft and Thagard, 2002) is an analogical mapper that uses visio-spatial representations, using the Java Visual Object System. It does no transfer of problem solutions and uses the ACME architecture for mapping (Holyoak and Thagard, 1997). MAGI (Ferguson, 1994) takes visual representations and uses the Structure-Mapping Engine (Falkenhainer et al., 1990) to find examples of symmetry and repetition in a single image through analogy. The Galatea system (Davies and Goel, 2001) uses only visio-spatial representations of problem-solving procedures and transfers a source solution to a target solution. By using a sufficiently abstract visual language it is able to transfer problem-solving procedures between semantically distant analogs. The work on Galatea also supports the notion that visio-spatial representations are useful for problem-solving.

Discussion

In this paper we described and demonstrated the applicability of the analogy methodology to the problem of secondary structure alignment from contact maps. Our hypothesis was that analogy using contact map similarity can effectively generate accurate protein substructure predictions. Triptych retrieves protein substructures based on visual similarity of contact maps. Initial results suggest that the retrieve and adapt phases are successful in finding similar contact maps in the PDB and modifying these to predict the alignment of pairs of helices, supporting this hypothesis. The advantage and novelty of our approach lies in its use of multiple sources of knowledge, including existing structural knowledge from the PDB, expert and text book knowledge (as used in the helix extension), as well as knowledge mined from the database. Once the viability of the approach is shown to be effective with idealized contact maps, the predicted, error-prone contact maps can be used as input.

Though not based on a cognitive model, Triptych shows how visio-spatial reasoning can facilitate problem solving in a another complex domain, building the case for the value of visual and spatial representations and reasoning for intelligent systems in general.

The theory behind this work is that changing representations can provide novel similarity insights. In this work we use contact maps and treat them as binary images, applying image processing techniques to them to retrieve similar protein substructures. In the adapt module, the information transferred is purely spatial. The success of this method for α-helix pair structure prediction provides preliminary support for this theory, in that generated visio-spatial representations can provide a means to find similarity. Future work will compare the results of contact map retrieval to sequence homology retrieval to investigate in exactly which conditions contact map similarity (representing visio-spatial representations) is superior to the non-visual homology similarity metric.

Acknowledgements

Funding provided by: the Natural Science and Engineering Research Council (Ottawa); Institute for Robotics and Intelligent Systems (Ottawa); Protein Engineering Network Center of Excellence (Edmonton).

References

Aaronson, J. S., Juergen, H., and Overton, G. C. (1993). Knowledge discovery in genbank. In Hunter, L., Searls, D., and Shavlik, J., editors, *Proceedings of the First International Conference on Intelligent Systems for Molecular Biology*, pages 3–11. AAAI Press.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000). Protein data bank. *Nucleic Acids Research*, 28:235–242.

Biname, J., Meurice, N., Leherte, L., Glasgow, J., Fortier, S., and Vercauteren, D. P. (2004). Use of electron density critical points as chemical function-based reduced representations of pharmacological ligands. *Journal of Chemical Information and Computer Science*, 44:1394–1401.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Alanysis and Machine Intelligence*, 8(6):769–798.

Casakin, H. and Goldschmidt, G. (1999). Expertise and the use of visual analogy: Implications for design education. *Design Studies*, 20:153–175.

Croft, D. and Thagard, P. (2002). Dynamic imagery: A computational model of motion and visual analogy. In Magnani, L. and Nersessian, N. J., editors, *Model-Based Reasoning: Science, Technology, & Values*, pages 259–274. Kluwer Academic: Plenum Publishers, New York.

Davies, J. and Goel, A. K. (2001). Visual analogy in problem solving. In Nebel, B., editor, *Proceedings of the International Joint Conference for Artificial Intelligence 2001*, pages 377–382, Seattle, WA. Morgan Kaufmann Publishers.

Falkenhainer, B., Forbus, K. D., and Gentner, D. (1990). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*, 41:1–63.

Faltings, B. and Sun, K. (1996). FAMING: supporting innovative mechanism shape design. *Computer-aided Design*, 28(3):207–216.

Farah, M. J. (1988). The neuropsychology of mental imagery: Converging evidence from brain-damaged and normal subjects. In Stiles-Davis, J., Kritchevsky, M., and Bellugi, U., editors, *Spatial Cognition– Brain bases and development*, pages 33–59. Erlbaum, Hillsdale, New Jersey.

Fariselli, P., Olmea, O., Valencia, A., and Casadio, R. (2001). Prediction of contact maps with neural networks and correlated mutations. *Protein Engineering*, 14(11):835–843.

Ferguson, R. W. (1994). Magi: Analogy-based encoding using regularity and symmetry. In Ram, A. and Eiselt, K., editors, *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, pages 283–288, Atlanta, GA. Lawrence Erlbaum Associates.

Gebhardt, F., Voss, A., Grather, W., and Schmidt-Belz, B. (1997). *Reasoning with Complex Cases*. Kluwer.

Glasgow, J. I., 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*, pages 53–62, Menlo Park, California. AAAI Press.

Gomes, P., Seco, N., Pereira, F. C., Paiva, P., Carreiro, P., Ferreira, J. L., and Bento, C. (2003). The importance of retrieval in creative design analogies. In Bento, C., Cardoso, A., and Gero, J., editors, *Creative Systems: Approaches to Creativity in AI and Cognitive Science. Workshop program in the Eighteenth International Joint Conference on Artificial Intelligence*, pages 37–45, Acapulco, Mexico.

Gonzalez, R. C. and Woods, R. E. (1992). *Digital Image Processing*. Addison-Wesley, New York.

Haralick, R., Shanmugam, K., and Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-3(6):610–621.

Holyoak, K. J. and Thagard, P. (1997). The analogical mind. *American Psychologist*, 52(1):35–44.

Jaccard, P. (1908). Nouvelles recherches sur la distribution florale. *Bulletin de la Société Vaudoise des Sciences Naturelles*, 44:223–270.

Jurisica, I. and Glasgow, J. I. (2000). Extending case-based reasoning by discovering and using image features in in-vitro fertilization. In *ACM Symposium on Application Computing (SAC 2000) Biomedical Computing- special session on biomedical applications of knowledge discovery and data mining*, Villa Olmo, Italy. CITO.

Jurisica, I. and Glasgow, J. I. (2004). Applications of case-based reasoning in molecular biology. *AI Magazine*, 25(1).

Monaghan, J. M. and Clement, J. (1999). Use of computer simulation to develop mental simulations for understanding relative motion concepts. *International Journal of Science Education*, 21(9):921–944.

Perner, P. (1998). Image analysis and classification of hep-2 cells in flourescent images. In *Fourteenth International Conference on Pattern Recognition*, pages 1677–1679, Brisbane, Australia. IEEE Computer Society Press.

Perner, P. (1999). An architecture for a cbr image segmentation system. *Journal of Engineering Application in Artificial Intelligence*, 12(6):749–759.

Perner, P., Gunther, T., and Perner, H. (2003). Airborne fungi identification by case-based reasoning. In *Workshop on CBR in the Health Sciences ICCBR03*, pages 73–79.

Smith, J. and Chang, S. (1994). Quad-tree segmentation for texture-based image query. *Proceedings of the second ACM international conference on Multimedia*, pages 279–286.

Sullivan, G. and Baker, R. (1994). Efficient quadtree coding of images and video. *IEEE Transactions on Image Processing*, 3(3):327–331.

Wang, J., Stieglitz, K. A., and Kantrowitz, E. R. (2005). Metal specificity is correlated with two crucial active site residues in escherichia coli alkaline phosphatase. *Biochemistry*, 14;44(23):8378–8386.

Won, C. S., Park, D. K., and Park, S. J. (2002). Efficient use of mpeg-7 edge histogram descriptor. *Electronics and Telecommunications Research Institute Journal*, 24:23.