SPATIAL MOTIFS IN DESIGN

JANICE GLASGOW
School of Computing, Queen's University
Kingston, Ontario, Canada
janice@cs.queensu.ca

AND

SUSAN L. EPSTEIN
Department of Computer Science
Hunter College and the Graduate Center
The City University of New York, New York, USA
susan.epstein@hunter.cuny.edu

AND

NATHALIE MEURICE, ANDY BECUE,
DANIEL P. VERCAUTEREN
Laboratoire de Physico-Chimie Informatique
Facultés Universitaires Notre-Dame de la Paix
Namur, Belgium
[Nathalie.Meruice,Andy.Becue,Daniel.Vercauteren]@fundp.ac.be

Abstract. Spatial design problems are rarely solved solely from first principles. Rather, previous designs or portions of designs are reused for novel problems. An important attribute of any computational approach to spatial design reuse is the ability to represent, abstract, transform and combine spatial patterns or motifs. This paper considers the issues and some possible solutions for reasoning with motifs. It also demonstrates how motifs can be applied in two diverse problem domains: park design and drug design.

1. Introduction

Rarely to people begin from first principles when contemplating a new design problem. On every scale, from cities to molecules, we reuse designs. Spatial design reuse involves the ability to abstract, transform and reassemble *spatial motifs* to address the goals of a new problem. A spatial motif can be defined as an abstraction over a set of recurring patterns observed in a dataset; it captures essential features shared by similar or related objects. In many problem domains, including design, there exist special regularities that permit motif abstraction and reuse. Such regularities may also exist across domains. Consider the spiral motif, one that occurs often in nature in forms such as hurricanes, seashells and alpha helices (see Figure 1 (a)). This motif has also be incorporated in architectural design (see Figure 1 (b)).

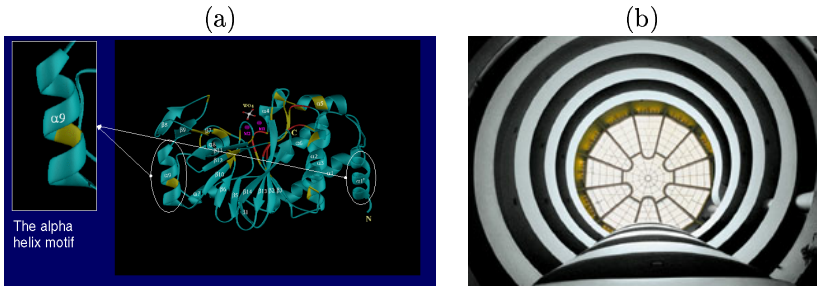


Figure 1: Examples of the spiral motif occurring in: (a) nature, in the form of an alpha helix in a protein structure, and in (b) architectural design, in the form of a six story spiral ramp in the Guggenheim museum.

Issues that need to be considered when applying motifs in a computational approach to design include:

- How are designs and motifs represented and transformed;
- How can motifs be abstracted from existing designs; and
- How are motifs combined to solve new problems.

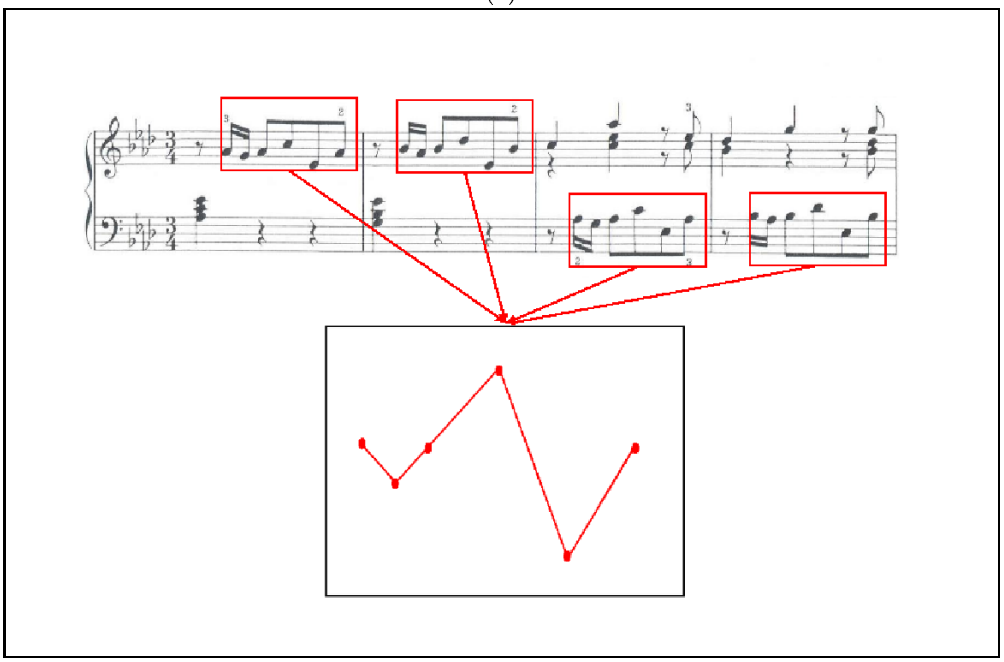
Figure 2 (a) illustrates the notion of motif representation and abstraction for a musical passage. The motif is represented as a two-dimensional graph that captures the order, relative pitch and duration of the notes in each of the passages. It does not preserve the precise pitch, thus several different passages can be abstracted to the same motif. Once a motif is abstracted it can be transformed then respecified. In Figure 2 (b) we abstract a motif from an initial passage, transform this through an inversion operation applied to the graph then respecify the motif as the new inverted passage.

In the remainder of the paper we explore in more detail the above issues involved in applying spatial motifs to design and some existing computational techniques that can be used to address them. We also demonstrate how using motifs can assist in two diverse problem domains: park design and drug design.

2. Motifs in Design

This section examines in more detail the issues of motif representation, abstraction and application to design problems. We do not attempt to propose universal solutions; rather, we present possible approaches for addressing the challenges of applying motifs in spatial design.

(a)



(b)

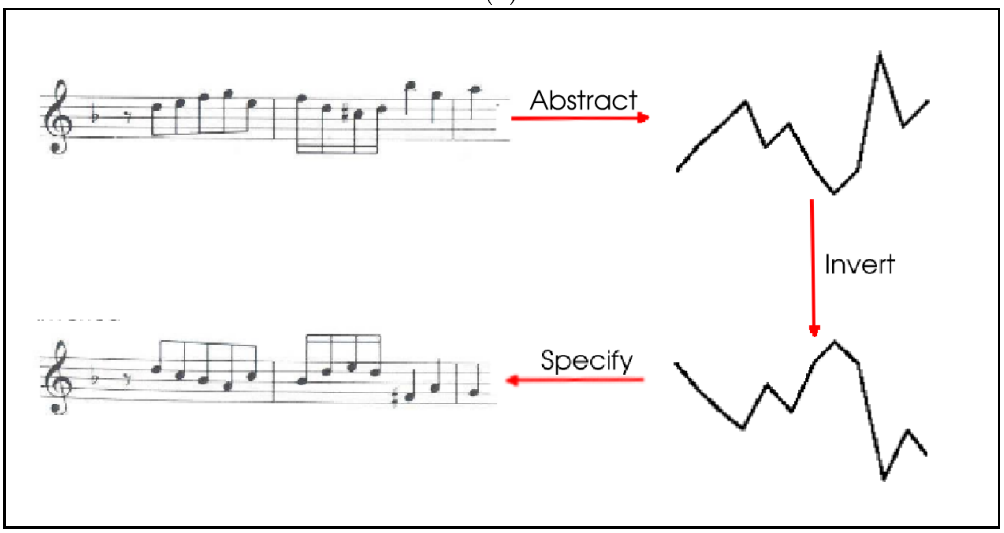


Figure 2: Examples of motif extraction (a) and inversion (b).

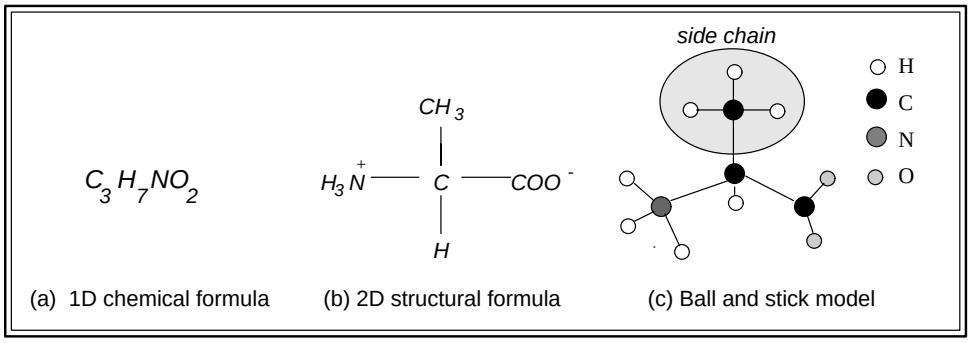


Figure 3: Alternative representations for a molecular structure.

2.1 REPRESENTATION AND TRANSFORMATION OF SPATIAL MOTIFS

The first question that must be addressed when considering a representation for spatial motifs is: What spatial relations need to be preserved in the representation? Consider the three representations for a simple molecule in Figure 3. The one-dimensional chemical formula (a) captures the chemical contents of the molecule. Additional information is available in the two-dimensional structural formula (b) that specifies connectivity among substructures. The ball and stick model (c) can be viewed as a three-dimensional representation that captures both the connectivity and relative location of the individual atoms in the molecule. The appropriate level of detail for a representation and the data structure we use to capture a motif is dependent on the questions we need to answer in our problem domain. For example, if connectivity among substructures is our primary concern then a structure such as illustrated in Figure 3 (b) may be most appropriate.

In terms of computation, it is not only important what features are to be extracted from the representation but also what transformations need to be applied in order to solve new design problems. Certainly a design and/or a motif can be captured and implemented in many forms: a logical description, a multi-dimensional graph, a symbolic array that preserves spatial locations, etc. However, it can be argued that a representation that explicitly preserves relevant spatial properties may be preferable in terms of criteria such as programming ease, efficiency and inferential adequacy (Glasgow 1993).

One approach to representing spatial objects and motifs that can be applied to design problems is computational imagery (Glasgow & Papadias 1992). This knowledge representation scheme, based on theories of mental imagery and model-based reasoning, incorporates three representations: spatial, visual and descriptive, each appropriate for a different kind of processing. It also includes a set of primitive functions defined to generate, transform and inspect image representations. For example, we can apply computational imagery to the blocks world to represent designs and motifs as multi-dimensional arrays and use the

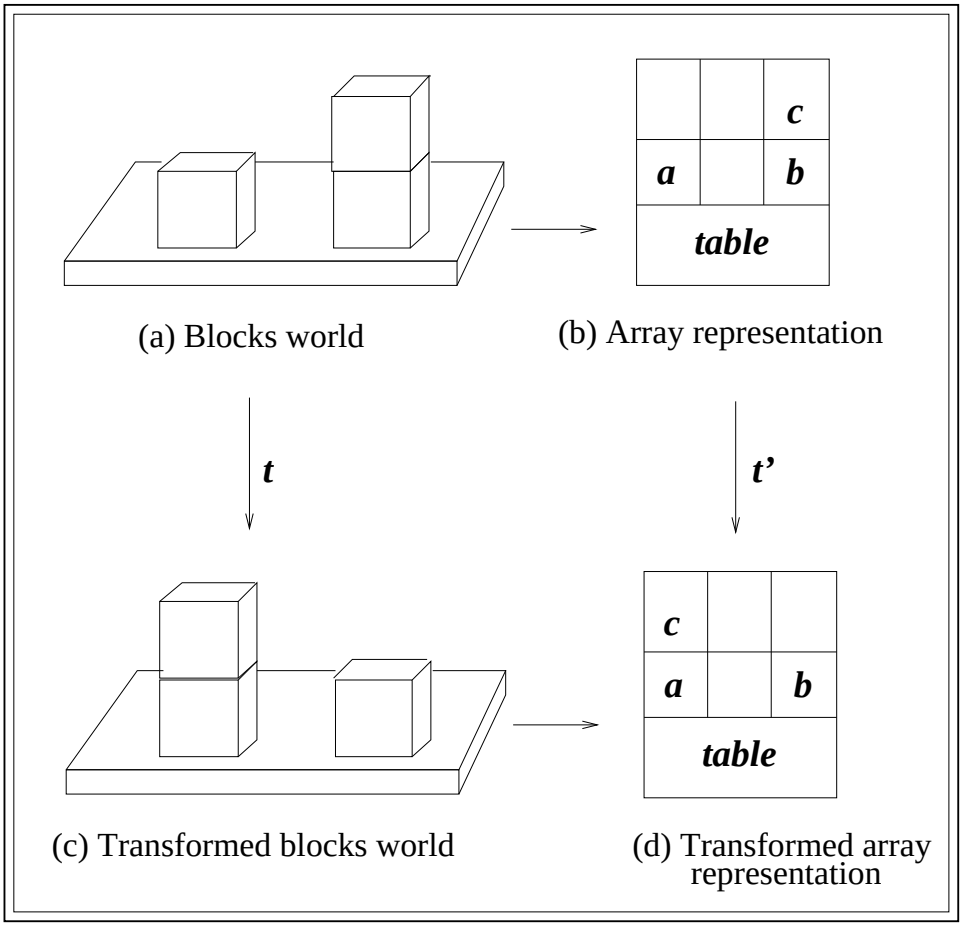


Figure 4: Application of a "move object c" transformation represented using arrays.

primitive operations (such as "move object" to transform a scene (see Figure 4)). Other image transformations include ones to resize parts of an image or motif, rotate a part and focus on details of a given part.

A representation may also be hierarchical in the sense that design may progress at different levels of complexity. As we will see in the park design application, initially we consider high level motifs (i.e., consisting of the main components of a park). Once the design is completed at this level, we then consider motifs for the lower level objects, such as play area, parking lot, etc. This is also true in the case of drug design where we consider representations of varying resolution and complexity.

Abstracting and respecifying objects in a design requires the representation of a conceptual hierarchy for objects in the domain. Take the problem of interior

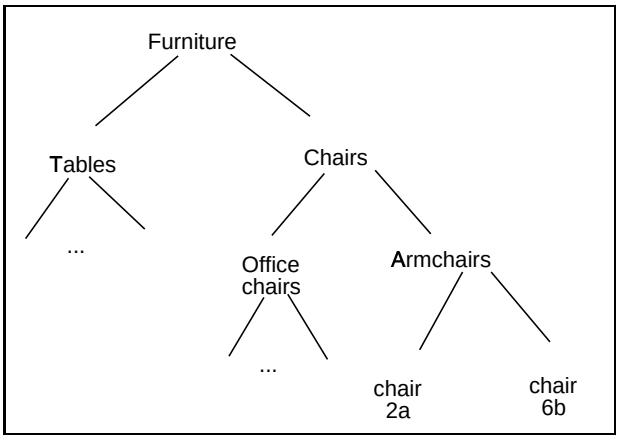


Figure 5: Conceptual hierarchy for objects in a design.

design. Figure 5 illustrates a conceptual hierarchy for possible furniture, where the leaves of the tree correspond to real objects which can be generalized and respecified during motif abstraction and specification.

2.2 MOTIF ABSTRACTION

Motifs may be designed by hand, based on expert knowledge. This is feasible when there are a limited number of previous, relevant designs to work from, as may be the case in a park design. In domains where motifs are extracted from a large experience base of designs, such as in drug design, automated knowledge discovery may be more appropriate. Knowledge discovery has been defined as "the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data" (Fayyad, Piatetsky-Shapiro & Smyth 1996). Generally, the automated discovery process is interactive and iterative, and can be broken into several steps (Brachman & Anand 1996): understanding of domain; creating a large data set; data cleaning and preprocessing; finding useful features to represent the data; data mining to search for patterns of interest; and interpreting and consolidating discovered patterns.

Many approaches to knowledge discovery and data mining have been proposed in the literature. However, most of these systems rely on an object representation that is expressed as a list of attribute-value pairs. Such a representation is not generally suitable for a spatial domain, where the most salient features of a design are the relationships among its parts. One method that has been developed for the discovery of spatial motifs is IMEM (Conklin, Fortier, Glasgow & Allen 1996). This conceptual clustering approach was specifically developed for the representation and classification of objects or scenes in terms of their parts and the relationships among these parts. These relationships may be topological (connectivity, proximity, nestedness) or spatial (direction, relative location, symmetry).

2.3 MOTIF COMPOSITION

Once motifs have been abstracted from previous designs, it is often necessary to reassemble them in novel configurations to address the goals of the new problem. One approach to this is by applying constraints that specify both preferred and prohibited configurations, as we will see in the park design domain. Heuristic search and/or constraint satisfaction techniques can be applied to find possible new designs. Once motifs have been configured then they can be specified to correspond to individual objects.

3. Applications

Following we present two applications, where motifs are used in the computational design of parks and of drugs. In both cases, previous designs are used to extract design patterns that can be reused in novel designs.

3.1 PARK DESIGN

The two-dimensional layout design for parks involves positioning a set of objects within a pre-specified outline (design frame) to meet a set of restrictions (criteria) (Epstein 1998, Epstein 2001). The purpose of a design frame is to anchor the design in a context. It distinguishes between the site, where the objects are to be placed, and the periphery, the site's border. A design frame includes relevant pre-existing objects, such as utility connections, transportation access, and bodies of water. An example of a design frame and a set of objects to be placed within it appears in Figure 6. The initial construction of a design frame may result from an abstraction over one or more previous designs. At the initial stage of design, the objects are abstract entities that are spatially configured to adhere to the given restrictions.

3.1.1 Representation

There are two kinds of criteria in a two-dimensional layout design problem: constraints, which must be satisfied, and principles, which are not required but are important in some unspecified combination. In every design problem, there is a constraint that prohibits any two objects from overlapping (e.g., no bench in a road). Other constraints describe each object as a set of property values that vary in degree of specificity and flexibility. A tennis court, for example, should have fixed dimensions, while a picnic area could vary more. Additional constraints may state how an object relates to the design frame (e.g., within 30 feet of the western boundary) or to another object (e.g., the tennis courts should be far from the pond). Other constraints describe the nature of the design frame (e.g., soil quality, slope, drainage) which could prevent the location of particular objects (e.g., no uphill tennis, no grass on clay-like soil). Principles, in contrast, are sensible or aesthetically desirable features (e.g., restrooms near the playground or a building with a good view of the pond, a picnic ground that

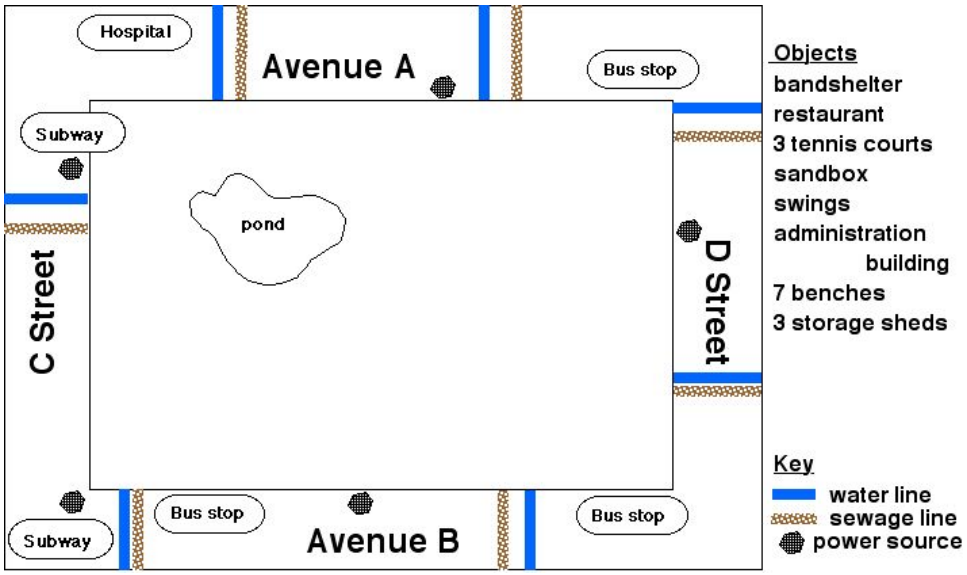


Figure 6: A design frame for a park problem

is not too steep).

A solution to a two-dimensional layout design problem anchors all the objects and satisfies all the constraints. To anchor an object, one must assign it fixed dimensions, a shape, a location, and an orientation. This corresponds to our notion of specifying a motif. Typically there are many solutions to such a problem; it is the mul-tiple, vague goal tests (as embodied in principles) that make such problems particularly difficult (Goel & Pirolli 1992). Although designers speak of solutions that are "better" or "worse", design is a search for a high-quality solution, not an optimal one (Goel and Pirolli 1992). The quality of a solution in two-dimensional layout design is measured by how well a solution serves all the tasks' principles.

Because design is so difficult, human designers are taught to work with varying levels of detail, beginning with functionally-specific use areas. Individual objects are targeted for particular use areas, but it is the areas that are placed within the design frame first. This is often done using motifs abstracted from previous designs of parks. When they turn to anchoring objects, human designers focus on crucial (major) objects (e.g., a restaurant) before less significant (minor) ones. Furthermore, they categorize objects by functional or structural similarities (e.g., benches, playing fields, roads) and treat them uniformly within category. Thus, there is a representational progression from the design frame plus the typed and targeted descriptions of the objects for the park problem (e.g., Table 1) to an area sketch (e.g., Figure 7), then to a plan that includes major objects (e.g., Figure 8), and eventually to a complete design.

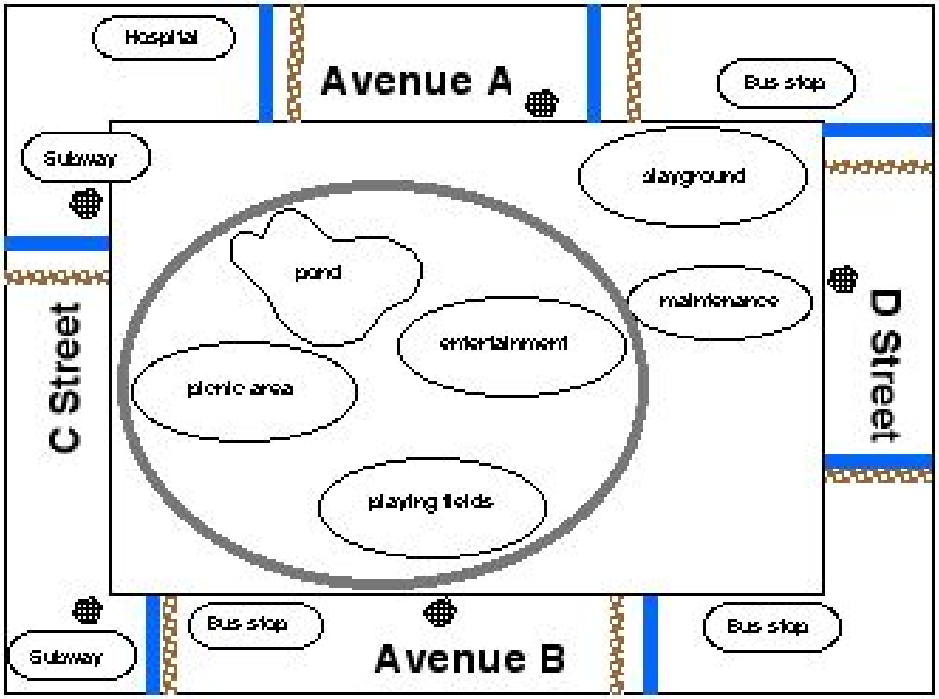


Figure 7: An area sketch with a road skeleton.

3.1.2 Park Motifs

A motif in two-dimensional layout design is a pattern that is repeated, either within the same solution, or from one solution to the next. For example, parks abound in particularly complex objects we call container networks, which are intended to provide transport or access (e.g., hiking trails or power lines). A container network is a skeleton (the principal arteries) plus branches (connectors to specific locations). The skeleton of each container network should appear in the area sketch, as the road skeleton does in Figure 7. Frederick Law Olmstead, thought of as the greatest urban park designer of the nineteenth century, often used a centrally-located ellipse as a road skeleton for urban parks (e.g., New York's Prospect Park and Central Park). An hour or two of experimentation with urban park layout engenders great appreciation for the elegance and simplicity of such a skeleton, or motif. In a particular park, there might be one branch from that skeleton to the restaurant.

A motif may occur at any level of representational detail. A second example is a collection of benches that form a seating cluster. A designer may experiment with many arrangements for such a cluster, and find different ones appropriate for different purposes. A conversation area may be U-shaped, benches for care

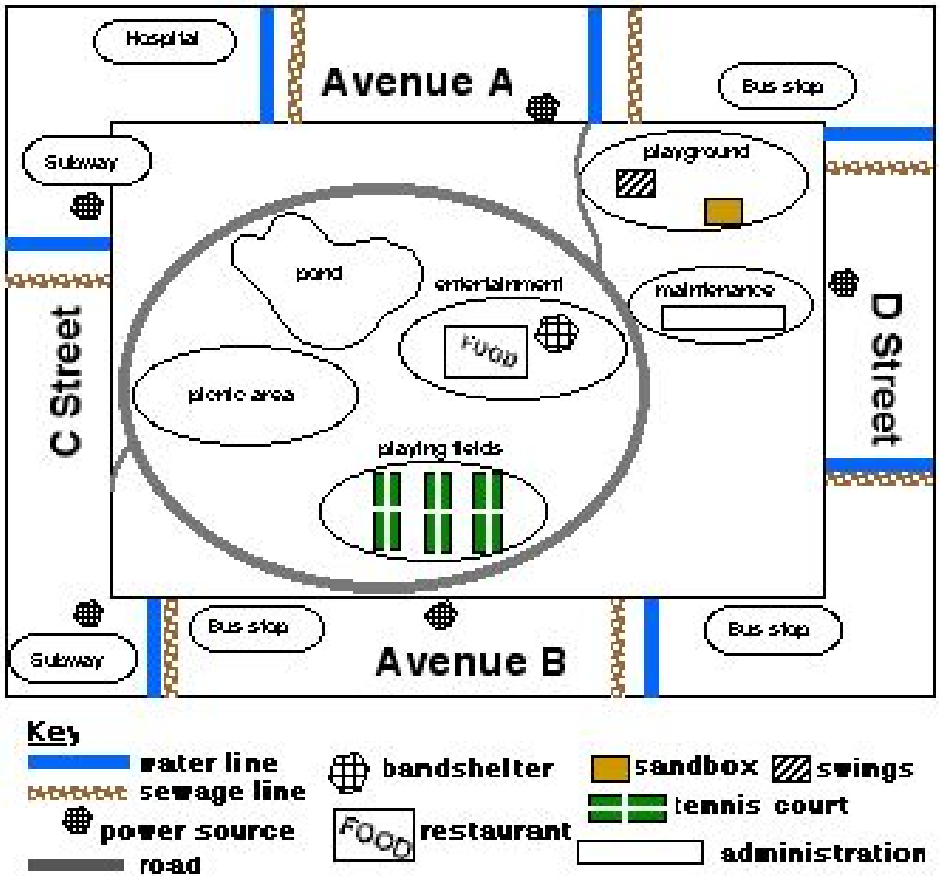


Figure 8: Area sketch extended to major objects and branches of the road network.

givers may circle a sandbox, and benches for pond-gazing may be strung out in an extended line. A third example is a collection of use area diagrams that relate three or more common use areas within a space (e.g., a playground, restrooms, and a pond). Using abstract entities and motifs to designate future decisions (e.g., "there will be seating here" or "the restrooms are near the playground") gives the designer leeway early in the process, when it is important to focus on the principles that can eventually lead to highly-valued designs.

To solve a new problem, a motif is subjected to a variety of transformations that have previously been specified as acceptable. An ellipse, for example, may be stretched in a variety of ways (but not so that two segments become very close to each other). A seating cluster may rotate slightly, or subdivide, or spread further, again within specified limits that prevent deformation from interference with function. Motifs can readily be combined in different areas of the same design (e.g., a playground-restroom-pond motif plus a playing-field-restaurant motif).

Motifs, and the constraints upon them, are, in theory, learnable. One might, for example, extract all of Olmstead's roads and make generalizations about their shape and/or relative location with respect to other entities. A greater challenge is to include more than one designer's products in the mix, and generalize from these. The result is a library of shared motifs (or cases), a set of patterns from which to create.

3.2 DRUG DESIGN

The rational design or discovery of novel compounds as interesting drug candidates for a specific biological target (receptor, enzyme, etc.) involves identifying the nature and spatial arrangement of the specific interactions required to observe significant and selective binding to the protein target, resulting in the appropriate biological activity. The motifs constituted by such structural requirements, as well as their topological relationships in space, are defined and reused by means of appropriate design strategies which strongly depend on the level of knowledge of the protein target.

When the three-dimensional structure of the host macromolecule is known, the motif of interactions governing the recognition phenomenon is directly identified by studying the nature of the amino-acids (see Figure 2 for an example of an amino acid) constituting the protein binding site, *i.e.,* the site in a protein is a particular pocket constituted by specific amino-acids buried in the three-dimensional structure made of alpha-helices, beta-turns and loops. Once established, the motif can be used either to screen huge libraries of molecules, or to design *de novo*, *i.e.*, build from scratch, new molecules that fit best both the shape of the binding site and the required specific interactions. In *de novo* approaches (Bohm 1995) small molecules or substructures are joined together to generate new structures. There are two major strategies for fragment joining. First, several fragments are positioned in the binding site in such a way to fit the motif of required specific interactions, and then search for an existing

molecule that connects the fragments into one molecule. Alternatively, you can start with a first element of the motif, Place a first fragment, and append additional ones in a step-by-step procedure. The second procedure tends to generate more flexible structures and allows the inclusion of chemical information such as synthetic feasibility.

In the difficult but common case where the three-dimensional structure of the protein is not available, as for transmembrane receptors, the direct study of the binding sites is not possible, and as a consequence, the evaluation of molecular similarity among a selected set of molecules (also called ligands) binding to the same protein target remains essential (Kuntz 1992). Such comparative approaches usually lead to models called "pharmacophore models" obtained by the best superimposition of substructures of the molecules (Wermuth & Langer 1993). These models, grouping the structural (and electronic) elements required for binding, as well as their spatial topology, can further be used to search for new drug candidates through database screening. The variety of chemical or pharmacological families, the number of ligands to consider in such comparisons, their flexibility, as well as their structural diversity however present severe difficulties that turn the superimposition process into a real challenge.

On one hand, for efficiency most existing molecular alignment methods are often specifically designed for a given family of pharmacological molecules, and are therefore rarely readily adapted to another pharmacological problem. On the other hand, such an approach is classically based on both the use of the nature and coordinates of the atoms constituting the molecules, as well as on several three-dimensional physico-chemical properties computed at an atomic level of resolution. The methods consequently fail to properly align compounds presenting similar binding abilities and pharmacological activities, but that are structurally significantly different, *i.e.*, constituted of very different chemical substructures, also called functions.

3.2.1 Representation

Approaches based on reduced representations describing the molecules at a chemical function resolution level (medium scale) rather than at an atomic resolution level (atomic scale), are particularly appropriate and powerful. They avoid the limitations described above and support the development of portable methods that can be applied to any set of biopharmacological molecules. More particularly, pharmacological ligands can be described by reduced representations based on the topological properties of their electron density, which gives the most accurate information on the stereo-electronic properties of the molecules. For example, in Figure 9, the three-dimensional structure of a given molecule is represented by its electron density (envelopes in middle images) at atomic level (upper part) and medium level (lower part). At atomic scale, the electron density envelopes correspond exactly to the atom positions, at medium scale, the three-dimensional structure is transformed into a more fuzzy representation, which interestingly can be further transformed, by topological analysis to an even more reduced representations constituted by selected relevant points of

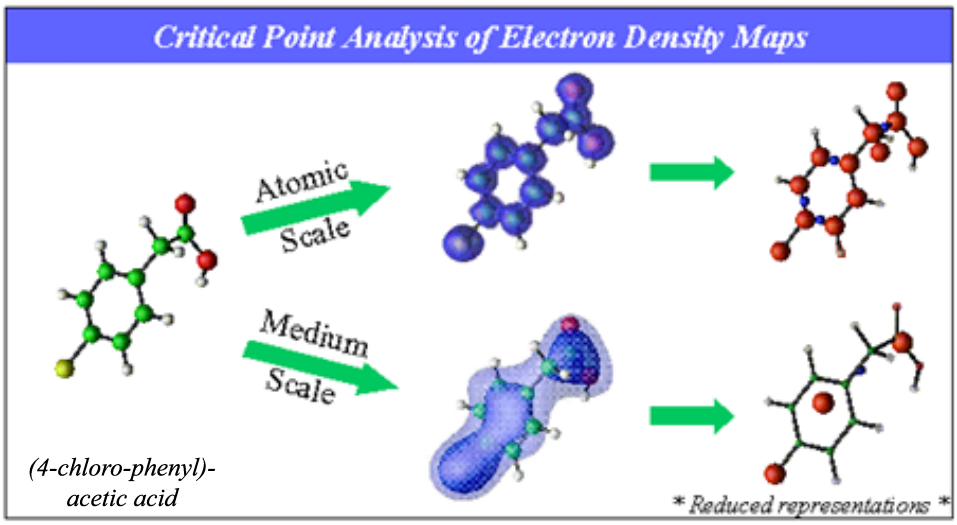


Figure 9: Views of the electron density (isosurfaces) and the corresponding reduced representations (spheres) of 4-chloro-phenyl-acetic acid at atomic and medium levels of resolution

the electron density, called critical points, which still contain all the important stereoelectronic characteristics. In the bottom right part of Figure 9, one sees that the three critical points at medium scale are sufficient to describe the functional groups of the molecules, *e.g.*, the chlorine atom (bottom), the phenyl ring (middle) and the carboxylic acid moiety (top).

3.2.2 Molecular Motifs

By assigning the critical points to the centre of the closest chemical function, any pharmacological molecule can thus be expressed in terms of a critical points pattern. Such critical points motifs can further be used either to design databases of descriptors directly correlated to the most common functions encountered in medicinal chemistry (Binamé, Meurice, Leherte & Vercauteren 2004), or to search for similarities among structurally dissimilar ligands. For example, Vercauteren and coworkers have shown that the critical points patterns describing molecules of interest may be superimposed (Meurice, Leherte & Vercauteren 1998, Leherte, Meurice & Vercauteren 2000, Meuricea, Rohrer, Maggiora & Vercauteren 1994); the resulting critical points alignments directly point out the matching chemical functions and can be easily interpreted in terms of molecular overlays leading to pharmacophore models.

Another advantage of critical point motif representation is the considerable reduction of the amount of data to handle. This occurs without loss of significant information. For this reason, such patterns are also useful in molecular complementarity applications and in particular for describing complex macro-

molecular systems such as proteins and DNA strings. At medium scale, it has been shown that the substructures of the proteins, as the alpha-helices represented in Figure 1(a), and the sub-parts of the DNA strings, as the sugars or base pairs, can be simplified to critical points motifs preserving the global three-dimensional topology as well as the stereoelectronic characteristics of the interacting partners (Becue, Meurice, Leherte & Vercauteren 2003). Such obtained motifs have been usefully implemented in an original docking tool to perform complementarity studies between two or three macromolecular systems (Becue, Meurice, Leherte & Vercauteren 2004). By designing such approaches, one will certainly help understanding how biomolecules, such as proteins and DNA strings, interact and recognize themselves, which is of major importance nowadays. Indeed, DNA and proteins are more and more considered as potential targets in biotechnology for the development of new efficient therapeutic agents.

4. Discussion

It is apparent that there is value in the reuse of old designs when considering a novel spatial design problem. To get value out of these involves abstracting patterns or motifs from the previous designs, which can then be transformed and reapplied to address the constraints of a novel problem. We have discussed some of the approaches for the representation of spatial motifs and designs, along with computational techniques for abstracting, transforming, combining and specifying motifs. Two diverse applications were considered, drug design and park design, that benefit from the use of design motifs.

We can conclude that a motif is an experience-rich, knowledge-rich reusable design segment. Some representation of a motif within a design engenders to coherence and harmony; too much may make the design non-unique and/or boring. Expert human designers typically have a vocabulary of design motifs, ones that can be reused in multiple projects. Future work in the area involves the incorporation of such expert knowledge as well as the development of automated techniques for extracting useful design motifs.

Acknowledgements

Funding for this research has been provided by the Natural Science and Engineering Research Council of Canada, the Institute for Robotics and Intelligent Systems (Canada), National Science Foundation (USA) IIS-0328743, PSC-CUNY, and the National Foundation for Scientific Research (Belgium).

References

Becue, A., Meurice, N., Leherte, L. & Vercauteren, D. (2003), 'Description of protein-dna complexes in terms of electron density topological features', *Acta Crystallographica* **D59**, 2150–2162.

Becue, A., Meurice, N., Leherte, L. & Vercauteren, D. (2004), 'Evaluation of the protein solvent accessible surface using reduced representations in terms of critical points of the electron density', *Journal of Computational Chemistry, submitted* .

Binamé, J., Meurice, N., Leherte, L. & Vercauteren, D. (2004), 'Use of electron density critical points as chemical function-based reduced representations of pharmacological ligands', *Journal of Chemical Information and Computer Science, accepted* .

Bohm, H. J. (1995), 'Site-directed structure generation by fragment-joining', *Perspectives in Drug Discovery and Design* **3**, 21–33.

Brachman, R. J. & Anand, T. (1996), The process of knowledge discovery in databases, *in* U. Fayyad, G. Piatetsky-Shapiro, P. Smyth & R. Uthurusamy, eds, 'Advances in Knowledge Discovery and Data Mining', AAAI Press/The MIT Press, pp. 37–57.

Conklin, D., Fortier, S., Glasgow, J. & Allen, F. (1996), 'Conformational analysis from crystallographic data using conceptual clustering', *Acta Crystallographica* **B52**, 535–549.

Epstein, S. (1998), Toward design as collaboration, *in* 'Proceeding of Fifteenth National Conference on Artificial Intelligence', AAAI Press.

Epstein, S. (2001), Toward autonomous layout design - an ai approach, *in* 'Proceedings of Visual and Spatial Reasoning in Design', Bellagio, IT, pp. 325–344.

Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996), From data mining to knowledge discovery: An overview, *in* U. Fayyad, G. Piatetsky-Shapiro, P. Smyth & R. Uthurusamy, eds, 'Advances in Knowledge Discovery and Data Mining', AAAI Press/The MIT Press, pp. 1–34.

Glasgow, J. (1993), 'The imagery debate revisited: A computational perspective', *Computational Intelligence* **9**(4), 309–333. Taking issue paper.

Glasgow, J. & Papadias, D. (1992), 'Computational imagery', *Cognitive Science* **16**(3), 355–394.

Goel, V. & Pirolli, P. (1992), 'The structure of design problem spaces', *Cognitive Science* **16**, 395–429.

Kuntz, I. (1992), 'Structure-based strategies for drug design and discovery', *Science* **257**, 1078–1082.

Leherte, L., Meurice, N. & Vercauteren, D. (2000), 'Critical point representations of electron density maps for the comparison of benzodiazepine-type ligands', *Journal of Chemical Information and Computer Science* **40**, 816–832.

Meurice, N., Leherte, L. & Vercauteren, D. (1998), 'Comparison of benzodiazepine-like compounds using topological analysis and genetic algorithms', *SAR and QSAR in Environmental Research* **8**, 195–232.

Meuricea, N., Rohrer, D., Maggiora, G. & Vercauteren, D. (1994), 'Molecular field and electron density-based comparative matching of benzodiazepine-like compounds', *Journal of Computer-Aided Molecular Design, submitted* .

Wermuth, C. & Langer, T. (1993), Pharmacophore identification, *in* E. Leiden, ed., '3D QSAR in Drug Design', pp. 117–136.

and the National Foundation for Scientific Research (Belgium).