On Scaffolds and Hopping in Medicinal Chemistry

Nathan Brown* and Edgar Jacoby

Novartis Institutes for BioMedical Research, CH-4002 Basel, Switzerland

Abstract: The molecular scaffold is an oft-cited concept in medicinal chemistry suggesting that the definition of what makes a scaffold is rigorous and objective. However, this is far from the case with the definition of a scaffold being highly dependent on the particular viewpoint of a given scientist. It follows, therefore, that the definition of scaffold hopping and, more importantly, the detection of what constitutes a scaffold hop, is also ill-defined and highly subjective.

Essentially, it is agreed that scaffolds should be substantially different from each other, although significantly similar to each other, to constitute a hop. In the latter, the scaffolds must permit a similar geometric arrangement of functional groups to permit the mode of action. However, this leaves the paradox of how to describe both scaffold similarity and dissimilarity simultaneously.

In this paper, the current statuses of scaffolds and scaffold hopping are reviewed based on published examples of scaffold hopping from the literature. An investigation of the degree to which it is possible to formulate a more rigorous definition of scaffolds and hopping in the context of molecular topologies is considered. These techniques are adapted from chemoinformatics to be applied in the design of new medicinal compounds.

Key Words: Scaffold, scaffold hopping, markush, isofunctional, lead hopping, chemoinformatics.

1. INTRODUCTION

It is of great importance in the development of Novel Chemical Entities (NCEs) to explore chemical series outside those that have been considered previously. This is important not only for investigating unprotected regions of chemistry space in terms of intellectual property, but also for opening up new lines of inquiry that have hitherto not been contemplated. This is crucial in diversifying our compounds in terms of chemical families that are returned initially from our High-Throughput Screening (HTS) campaigns [1].

In recent years, the term scaffold has been used extensively in the context of molecules generally to describe the central component of a molecule. However, although a number of scaffold definitions and determination algorithms have been mooted, no single approach covers all of the features that are intuitively important to all practitioners for which it is required.

The determination of a scaffold is particularly important in identifying new chemical series in a process referred to as scaffold hopping (also referred to in the literature as lead hopping, leapfrogging, chemotype switching, and scaffold searching) [2]-[5]. Therefore, if a suitable definition for a scaffold cannot be settled upon agreeably, then the definition of what constitutes a scaffold hop remains equally difficult to formulate objectively and reliably.

In this review, consideration is given to the variability of definitions of scaffolds and how this affects the follow-on challenge of scaffold hopping. The paper concludes with a review of methods published recently to effect scaffold hopping using topological representations, where no knowledge is assumed of a binding site or bound conformation.

*Address correspondence to this author at the Novartis Institutes for BioMedical Research, CH-4002 Basel, Switzerland; Tel: +41 (0) 61 324 90 29; Fax: +41 (0) 61 324 33 57; E-mail: nathan.brown@novartis.com

2. MOLECULAR SCAFFOLDS

The term scaffold is now used extensively to describe the core structure of a molecule. Taken literally, the core structure is the central component of a molecule: the substantial substructure that contains the molecular material necessary to ensure that the functional groups are in a desired geometric arrangement. However, the core structure can also simply refer to the key component or components of the molecule that a particular scientist defines, and not necessarily a scaffold in the literal sense.

Experts with different backgrounds and knowledge will tend to define a scaffold differently depending on their particular domain of interest. For instance, a medicinal chemist may define a molecular scaffold based on the diversity not of their perception of a core structure, but on the relative diversity of the synthetic routes to the molecules themselves. Whereas, patent lawyers would typically consider only the general similarity of the internal structure of the molecule to determine whether or not that particular region of 'scaffold' chemistry space has prior art in terms of its Markush structure representation (see below) for the particular application domain of interest. Chemoinformaticians, however, will always favor an objective and invariant algorithm that will provide a solution rapidly and without ambiguity. In this case, a scaffold definition is provided by a graph transformation algorithm that, given a molecular topological graph, ensures that the scaffold can be realized deterministically. However, there are also significant limitations in the scaffold determination algorithm that maintains current favor in Chemoinformatics.

1389-5575/06 $50.00+.00

2.1. Early Definitions of Scaffolds

One of the earliest scaffold descriptions was that introduced by Eugene A. Markush of the Pharma-Chemical Corporation in a landmark patent that was granted on 26th August, 1924 [6] – although this was not the first patent to include such a generic definition. Markush's patent covered an entire family of pyrazolone dye molecules:

The process for the manufacture of dyes which comprises coupling with a halogen-substituted pyrazolone, a diazotized unsulphonated material selected from the group consisting of aniline, homologues of aniline and halogen substitution products of aniline.

In making this claim, Markush was able to claim rights to not just an individual compound of interest, but also a large number of molecules of only potential interest in the chemistry space surrounding the actual molecule synthesized at the centre of the claim. Markush structures are now used extensively to protect chemical series of interest in patents in any industry that develops NCEs. Since the Markush generic structure has an intellectual property rights, rather than a scientific, founding it is not of absolute necessity that the molecules are chemical realizable.

The previous scaffold definition is for intellectual property applications, but the scientific definition is also important to describe accurately and invariantly. However, the definition of a scaffold is deceptively trivial to state, but incredibly difficult – if at all possible – to reduce to a set of generic rules that do not consider how the definition will be applied. For an early reference for an acceptable definition of a scaffold, as we tend to mean it today, we can go to the American Chemical Society (ACS) literature database and a paper from 1969 [7] that describes it thus:

The ring system is highly rigid, and can act as a scaffold for placing functional groups in set geometric relationships to one another for systematic studies of transannular and multiple functional group effects on physical and chemical properties.

Although the definition is clear in explanation it does not provide the practitioner with a rigorous and invariant description to allow the determination of the scaffold component of any given molecule.

2.2. Medicinal Chemistry and Molecular Scaffolds

Historically, medicinal chemists have used intuition to design molecules that exhibit a particular biological activity and that are sufficiently structurally dissimilar to extant molecules so as to be defined as a new chemical class. As a result it is difficult if not impossible to define a set of rules or heuristics that were used since each approach was influenced largely by the biological activity being sought rather than a particular set of defined rules that can be applied in general. The necessity was to develop molecules that are isofunctional but are sufficiently different from extant molecules so as to be novel: mode of action, selectivity, potency, other further aspects of activity, administration, availability at the target site, or other metabolic properties. This process included not only bioisosteric replacement, but also significant alterations and substitutions to the underlying structure to bring about both novelty and the desired response.

2.3. Recent Growth of Occurrence of Scaffold Terms

Over recent years there has been a marked increase in the use of the term scaffold when considering molecular classes or chemotypes – and even more recently the concept of scaffold hopping, which will be considered in the following section of this review. Much of this interest has arisen through the advent of combinatorial chemistry, and this is the most likely reason why our current methods tend to emphasize some core structural component as the molecular scaffold.

We can observe this substantial growth if we consider a series of text searches of the entire ACS literature catalogue by year (http://pubs.acs.org). By querying the occurrences of the terms 'scaffold' and 'scaffold hopping' in each of the titles, abstracts and main text, respectively, of the articles in the archive it is possible to develop an understanding of the perceived rate of growth of the use of these terms in the field of chemistry. It is accepted that these terms are searched in a general context, but we feel that the results should provide at least a good indicator of the growth of usage of these terms – with an anticipated similar trend with the various syno-nyms of scaffold – in the field of chemistry. The results are reported for each year as the proportion of the entire set of articles that contain these terms.

From Fig. (**1**) it can be observed that only in the last 10 years or so (1995 to 2005) has there been significant growth in the use of the 'scaffold' term – prior to 1980 the results were negligible or non-existent. While, in Fig. (**2**), the use of the 'scaffold hopping' term has only really been evident from 2000 onwards, with 50 papers published that contained the scaffold hopping phrase in 2005. Part, and perhaps a substantial amount, of the increase in the use can be attributed to the advent of combinatorial chemistry. However, it is also difficult to separate that field with the field under consideration here. It can also be generally observed that there has indeed been an increase in the interest in scaffolds and scaffold hopping in general over recent times.

2.4. Computational Scaffold Abstractions

None of the references above provides us with an invariant computational method for determining the scaffold component of a molecule, which is essential for the nature of our current efforts in deriving information and understanding from our vast chemical libraries. In 1996, Bemis and Murcko [8], proposed the molecular and graph frameworks of molecules as invariant approaches to determine scaffold representations of molecules such that the resultant scaffolds could then be applied to classification problems and the objective determination of scaffold hopping.

From a single molecule, it is possible to generate the Bemis and Murcko scaffold or molecular framework, as well as the graph framework, as required. The former prunes sidechains of the molecule, but maintains the original atom typing and bond orders used. The latter takes the same molecular framework and then proceeds to further abstract the atoms and bonds to uncolored and unweighted nodes and

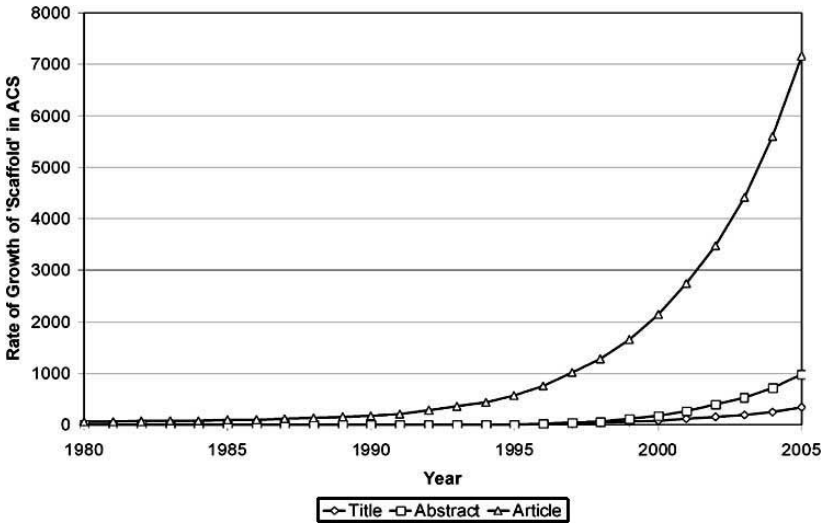


Fig. (1). The cumulative rate of growth of the use of the term 'scaffold' in the title, abstract, and articles from the ACS literature archive in the last 25 years (1980 – 2005); pre-1980 figures are negligible.

edges, respectively, thus giving an indication of the general framework of each molecule considered. The graph frameworks can be further abstracted by representing the ring systems as nodes of the graph. An example of a molecule, the anti-hypertensive drug Diovan® (Fig. **3a**) with its molecular, graph, and reduced scaffold frameworks is provided in Figs. (**3b**, **3c** and **3d)**, respectively.

The scaffold frameworks defined here have found great application in reducing the over-representation of particular chemical classes in subset problems where the aim is to rationalize the sizes of the chemical series in lead-finding exercises [1]. Additionally, the methods have been applied as an approach to determine whether or not – and the degree to which – a scaffold hop has been achieved. However, the algorithm can suffer from pathological conditions in a number of cases. In instances where a molecule is acyclic then, according to the Bemis and Murcko rules, there is no definable scaffold. This therefore leads to null scaffolds. Furthermore, additional consideration is necessary to determine whether or not a particular molecular scaffold is a substructure of another molecular scaffold. This is a particular challenge in situations where substituents are themselves ring systems, since they will be contained as part of the scaffold and therefore different to the 'true' scaffold. In cases such as these, it is important to detect these limitations to ensure improved application of the Bemis and Murcko method. Recent efforts have been made to overcome this latter issue by defining scaffold trees that indicate the inheritance traits of scaffold families [9-11] and this approach has been demonstrated to be applicable to the generally more complex natural products databases [12].

One particular method of defining scaffolds is to cluster a given dataset based on structure into defined classes and then perform a Maximum Common Substructure (MCS) search

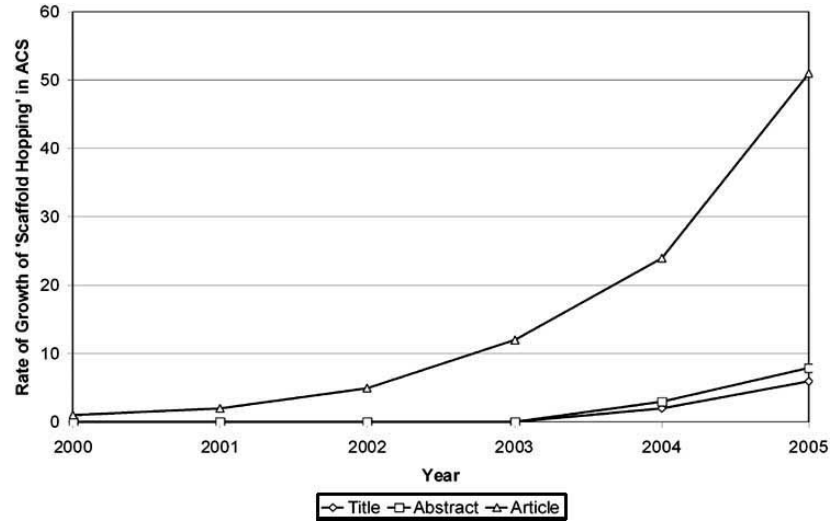


Fig. (2). The cumulative rate of growth of the use of the term 'scaffold hopping' in the title, abstract, and articles from the ACS literature archive in the last 5 years (2000 – 2005); pre-2000 figures are negligible.

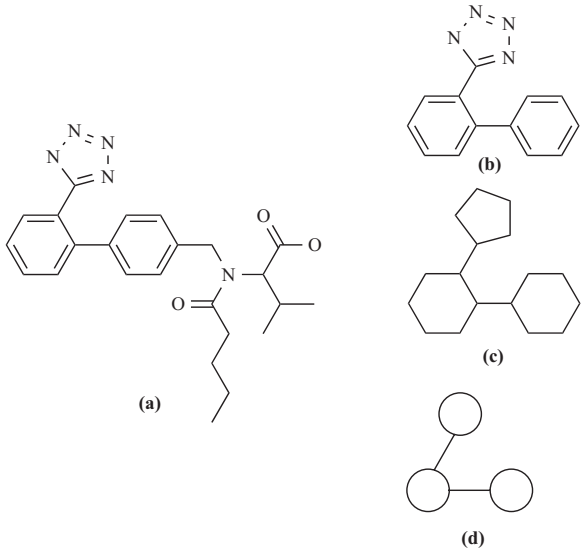


Fig. (3). The (b) molecular, (c) graph, (d) reduced scaffold framework representations for the anti-hypertensive Diovan® molecule (a), respectively.

for each of these classes to determine the MCS for each cluster [10]. This provides a more naturalistic method of scaffold detection by grouping similar objects and discovering the largest degree of structural similarity they have in common. However, a drawback to this approach is that the method is dataset dependent, since different datasets will result in alternative clustering results and, subsequently a different MCS for the clusters.

Regardless of these limitations, the Bemis and Murcko scaffold frameworks have been adopted widely in the field as is evident by the literature. The approach is by no means suitable in all cases, but it does provide us with a method that can be applied generally in our analyses while also providing a reasonably effective indicator of whether or not a scaffold hop has occurred.

The definition of a scaffold is by no means a solved problem and is prone to a wide range of subjective viewpoints that are necessitated by the particular problem domain under consideration. However, the general acceptance of the Bemis and Murcko methods for scaffold representation determination in computational analyses has permitted a number of structural analyses that allow us as scientists to select molecules based on their likely chemical class in an unbiased and repeatable manner and also to suggest whether we have achieved the goal of jumping from one chemical series to another.

3. LIGAND-BASED SCAFFOLD HOPPING

Scaffold hopping has quickly become a very important area of investigation in drug discovery to determine or design novel molecular backbones that still permit the desired response to be observed when the backbone is decorated with substituents to evoke that function [5].

When a binding site or a pharmacophoric model is known then it is possible to use *de novo* design methods [13,14] that are to develop novel molecules with significantly different scaffolds thereby providing potential areas of greater potency and novel intellectual property. Extant molecules may also be virtually docked against the protein binding site to investigate the potential binding energies of the molecules [15-17]. However, when the receptor structure information is not known, it is necessary for extant molecules that are known to invoke a desired response to be applied in an attempt to discover or develop new molecules with diverse scaffolds that maintain a likelihood of being of interest in terms of the objective of the study: ligand-based scaffold hopping.

The availability of only topological information necessitates the development and application of molecular descriptors that both emphasize the similarities that are considered important for the response being measured in an invariant way, while also providing an abstracted description of what is largely unimportant for the same response. Therefore, to over-simplify somewhat, the search is for something that is similar in response but different in structure. However, the topological and geometric arrangements of the potential interaction points of a molecule are also highly important and must be taken into account to some degree to ensure that the newly discovered or designed molecule will retain a suitable geometric similarity to the query molecule in terms of its perceived key pharmacophoric points. This can be differentiated somewhat from scaffold replacement strategies where the aim is to identify new scaffolds to replace an extant scaffold that will still maintain the spatial arrangement of the actual substituents in the molecule, rather than potential interaction points.

The challenges in the discovery or design of molecules with novel scaffolds for a particular response of interest can be summarized with the matrix in Fig. (**4**), where the dimensions consider the similarities in a particular property and

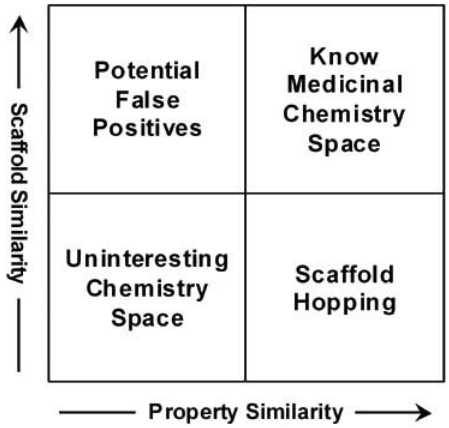


Fig. (4). Illustration of the objectives in scaffold hopping where the aim is to locate an isofunctional molecule that has a scaffold – however defined – that is significantly different from our known structures; the lower-right quadrant. Traditional similarity searching only returns molecules from the upper-right quadrant and, although interesting in a wide range of circumstances, does not provide us with information about truly novel regions of chemical space.

scaffold space, respectively. According to this matrix, the aim in scaffold hopping is to discover or design compounds in the lower-right quadrant where the outcome is similarity in our desired property space, but dissimilarity in our specified (and as mentioned previously, potentially arbitrary) scaffold space. In the more traditional connection-based similarity searching approaches, molecules from the upper-right quadrant are typically returned, essentially molecules from a region of chemistry space that has likely been explored widely in previous studies. Our aim in scaffold hopping is to move away from what is in the known region of chemistry space towards new regions of chemistry space that will result in molecules with similar properties yet diverse frameworks.

Therefore, in situations where only the topology of a molecule of interest is known, it is necessary to employ structure representation methods that do not explicitly codify, or at the very least de-emphasize these structural features, the connection or topological information within the molecule, but instead attempts to obfuscate this information permitting the search process to be relaxed and consider only those aspects of the molecule of interest that are expected to be of significance, namely the functional groups and their geometric orientations. What follows is a brief review of recent methods that have been published for ligand-based scaffold hopping.

3.1. Objective Measures of Scaffold Hopping

Many approaches have been posited for scaffold hopping *in silico*; however the validation of these approaches has quite often been based upon visual inspection of the returned structures and therefore highly subjective and variable. However, for a scientific study it is necessary to define objective measures for determining reliable indicators of scaffold hopping. Recently, efforts have been made to provide objective validation protocols in retrospective studies to determine with some level of confidence whether the results of these methods are likely to be occurrences of scaffold hopping.

The chemotype hit rate was proposed by Good *et al.* [18] as a method of investigating not simply the overall enrichment of actives in a virtual screening campaign, but also the number of chemotypes, however defined, that are recalled with the approach. In this approach, the chemotypes appear to have been defined based on structural moieties of recalled structures rather than a Bemis and Murcko approach. In a similar approach, Stiefl *et al.* [19], recently proposed a method of measuring objectively both the recall of actives, while also giving due consideration, where appropriate, to the chemical classes that are also covered in the recalled compounds, again applying visual inspection together with clustering methods to both subjectively and objectively investigate the quality of the results in terms of scaffold hopping.

Although the limitations of the Bemis and Murcko scaffolds have been noted already, this scaffold determination method does provide a method of investigating the extent to which a newly found ligand or set of ligands from a scaffold hopping campaign constitutes a scaffold hop. By reducing the query ligand and the retrieved ligands to the molecular and graph frameworks, respectively, a comparison may be made to determine the extent to which a scaffold hop has been achieved. This comparison may be made either by visual inspection or, more objectively, through molecular similarity methods using topological or whole-structure descriptors to determine a quantitative degree of similarity.

The MEQI (Molecular EQuivalence Indices) keys [20] define a one-way transform from molecular structure or substructure to a 5 alphanumeric character key, or *meqnum*. The MEQI approach may be used in scaffold hopping to evaluate the chemotype class membership by generating the keys for each of the possible abstractions of the molecules: whole molecule representations; Bemis and Murcko frameworks with no abstraction, atom, atom and bond, and ring abstractions, respectively; and also the constituent cyclic fragments separately as cycle size and abstracted cycles, respectively. An illustration of meqnum key generation is provided in Fig. (**5**) providing keys that represent the entire structure and two subcomponents.

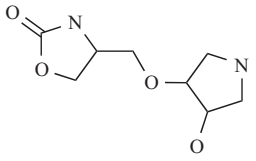
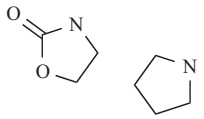
	meqnum		Structure
(a)	A9P8B		
(b)	KYYT5	NRR57	

Fig. (5). The unitary meqnum (a) of the whole structure with the ring systems taken jointly and (b) its two composite ring system meqnum representations.

A further approach recently proposed by Hert *et al.* [21] calculates the Mean Pairwise Similarity (MPS) of the set of returned structures with the expectation that an instance of scaffold hopping will be apparent in instances when the returned set has a lower MPS value. This method of measuring structural diversity indicates the overall structural diversity covered by the dataset that is under analysis. Again, the descriptors to calculate the MPS should be those that provide a whole-structure description, thereby explicitly encoding the internal structure – and therefore its scaffold information – of the molecule. However, the set of retrieved structures may contain a large number of close analogues (dependent on the similarity method applied) and it therefore may be prudent to rationalize the retrieved set by removing those structures that contain an identical or similar Bemis and Murcko scaffold to the query molecule.

A related approach considers the structure of the data of a dataset known *a priori* that permits an estimation of Tanimoto coefficient values or ranges at which scaffold hopping characteristics may be expected [22]. The approach uses the

Tanimoto values of the pairwise comparisons of all of the molecules in a given dataset from which the distribution of Tanimoto similarities can be analyzed to indicate a likely lower-bound Tanimoto similarity below which scaffold hopping would be more likely to occur. The authors of this study concluded that it was of great importance to consider the structure of your given data itself rather than rely on a potentially dataset dependent heuristic Tanimoto cuts-off that have been proposed to suggest a confidence limit for similar biological activity to be present.

Given the vagaries surrounding what constitutes a molecular scaffold, it is advisable that a selection of these objective measures be used to provide a greater assurance as to whether a scaffold hop has indeed been achieved.

4. MOLECULAR DESCRIPTORS FOR LIGAND-BASED SCAFFOLD HOPPING

The specificity of a molecular descriptor method is important in scaffold determination and its applications. Generally, it is considered that a molecular descriptor should be specific: that is, the description should be as unique as is possible so that a molecule maps to a unique region of our descriptor space. However, in the context of scaffolds and scaffold hopping, it tends to be more important to blur our descriptions such that molecules with the characteristics in which we are interested being mapped to the same or similar regions of our descriptor space, there by allowing us to readily access these structures.

Currently, three general approaches have gained currency in ligand-based scaffold hopping: topological, geometric and surface-based (or field-based), representations. Each of these methods characterize the given structure in an abstracted form thereby altering the rigid connection-based or structural chemistry space in which we are searching to reflect the particular application for which the methods are intended.

Abstraction of molecules for the purpose of scaffold hopping should focus on the information-rich components, whether these are substructures or inter-relationships between molecular features – where features can be defined as atoms and/or substructures. Typically, the most information-rich components in a molecular tend to be heteroatoms and the atoms at the extremities of the molecule. In the topological descriptor methodologies described here, the approaches tend to describe the most information-rich atoms – as in topology-based methods – or to find a lower dimension projection of the intramolecular space that maximizes the information content in the chosen projection – as in topography and surface-based methods.

4.1. The CATS Family

The Schneider laboratory has published a family of three ligand-based scaffold hopping descriptors known collectively as the CATS (Chemically Advanced Template Search) descriptors: CATS, CATS3D, and SURFCATS, respectively. Each member of the family takes advantage of different representations of the molecules under consideration: topology, topography, and surface, respectively. These descriptors are classified generally as Correlation Vector Representations (CVR), and were first reported in the 1980s [24]. Here, a brief overview is given of each of the descriptor methods. A diagrammatic summary of each is provided in Fig. (**6**).

CATS

Perhaps the first molecular descriptor that was defined explicitly for the challenge of scaffold hopping were the

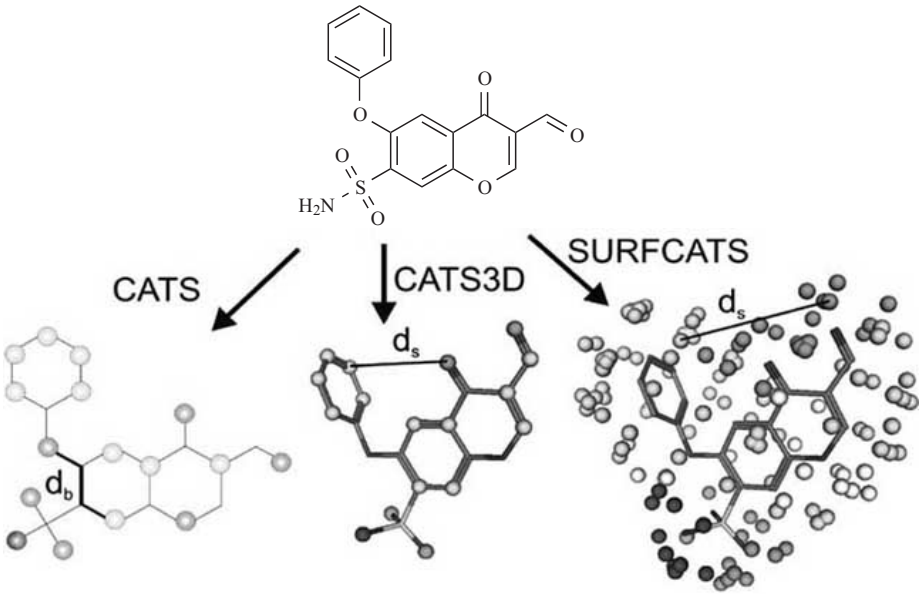


Fig. (6). The CATS family of descriptors: CATS, CATS 3D, and SURFCATS, respectively. Image courtesy of Renner and Schneider [23] and reprinted here with permission.

CATS two-dimensional (2D) vectors from Schneider *et al.* [2]. The method proceeds initially by abstracting each of the nodes in a given molecular graph to one or more of the following atom types as in Fig. (**7**): hydrogen-bond **D**onor, hydrogen-bond **A**cceptor, **P**ositively charged, **N**egatively charged, and **L**ipophilic. Nodes that cannot be classified as one of the above abstracted atom types are not assigned a type, or remain uncolored.

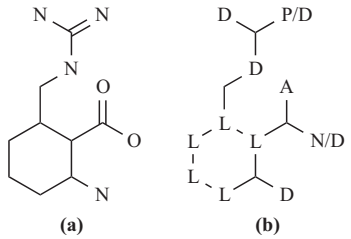


Fig. (7). Example of the conversion of a molecular graph (a) into the generalized CATS molecular graph. The figure has been redrawn from Schneider *et al.* [2].

The algorithm then proceeds by calculating the shortest graph edge path between all abstracted node pairs, bar the remaining uncolored nodes, up to a maximum distance of 10 bonds. Since, from the 5 generalized atom types described above, there are 15 unique atom pairings and only graph edge distances from 1 through 10 are considered, this results in 150 unique vector positions, with each position corresponding to the frequency of occurrence for each abstracted node pair at a given graph distance.

CATS3D

CATS three-dimensional (3D) vectors are an extension of the standard CATS vectors that uses the Euclidean distance between the atom in the molecule in ångströms (Å) [25]. A single conformer is generated using the CORINA software [26]. The generalized atom types – assigned to all atoms, including hydrogens – in CATS3D are a set of seven features: cationic, anionic, polar, acceptor, donor, hydrophobic, or other. Each of these are from the set of types defined by the PATTY (Programmable ATom TYper) typing scheme [27]. This scheme results in 28 unique atom pairs and 20 distance bins (from 0 to 20 Å) resulting in a vector of length 560.

SURFCATS

The SURFCATS descriptors, also from the Schneider laboratory, were defined recently [23] and essentially turn the CATS and CATS3D descriptors inside out and concentrate on describing the key surface points of the molecule under consideration. The same 20 distance bins as for CATS3D were used. The surface of a molecule is realized by the Gauss-Connolly function from the MOE software [28] with a spacing of 2 Å.

The CATS family of descriptors was published recently in a comparative study [23] for scaffold hopping where it was discovered that the application of all members of the family provides disjoint sets of solutions. The recommendation from this result, therefore, is that a range of possible molecular representations can be applied usefully in real-world situations.

Two further descriptors have also been proposed by the Schneider laboratory [29] for scaffold hopping and these are discussed here in brief.

CHARGE3D

The CHARGE3D descriptor vector representation [24] uses the partial atomic charges as atomic descriptors, with each atom pair – including hydrogens – being specified by the product of its two partial atomic charges together with the Euclidean distance between those two atoms in a single conformer generated by the CORINA software. The Euclidean distances are binned into 100 individual positions on a vector from 0 to 10 Å in 0.1 Å increments, with all distances greater than 10 Å added to the final bin in the vector. The vector positions are incremented with the relevant product of the atom pair partial charges for the particular vector position that corresponds to the Euclidean distance between the two atoms.

SQUID

The SQUID (Sophisticated QUantification of Interaction Distributions) descriptors again use pharmacophoric abstractions of the atoms as with the CATS3D vectors [30]. However, SQUID uses alignments of multiple ligands of interest to define a weighted pharmacophore model with preserved features gaining a higher significance in the pharmacophore model using spatial Gaussian probability densities. All the pairs of these Potential Pharmacophoric Point (PPP) distributions are then taken together with the interatomic distance in ångströms. These PPP pairs are then encoded into an identically formatted vector as for the CATS3D vectors.

4.2. Similog

Similog pharmacophoric keys are similar to the topological CATS variant in that the nodes of a molecular graph are generalized by four features with these being represented as four bits per atom. However, Similog keys represent triplets of atoms rather than pairs, thereby increasing the size of the potential key space [31]. In this way they can be seen as topological analogs of traditional 3-point geometric pharmacophore models.

The keys are based on a DABE atom typing scheme in which the constituents are: potential hydrogen bond **D**onor; potential hydrogen bond **A**cceptor; **B**ulkiness; and **E**lectropositivity. Therefore, theoretically there are 2^4 potential atom types; however, 6 of these are not possible with neutral organic species and the key 0000 is uninformative, leaving 9 possible keys. All possible triplet combinations of these 9 keys give a theoretical limit of 8031 triplets; however, only 5989 had been found at the time the original paper was written [31]. The sorted triplet keys are each encoded into a binary vector of length 5989, to denote presence or absence of a particular key. A summary of Similog key construction is given in Fig. (**8**).

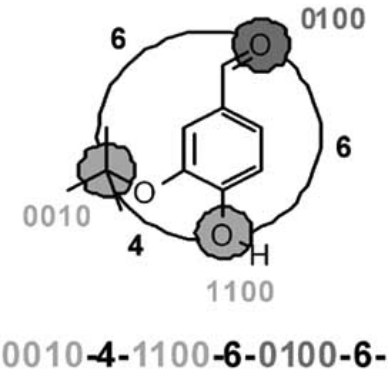


Fig. (8). Example of a Similog key determination. The figure has been redrawn from Schuffenhauer *et al.* [31].

4.3. Reduced Graphs

Another approach that has been employed to develop generalized or abstracted representations of molecules is the reduced graph (RG) representation [32,33]. In this method, entire substructures, according to some defined scheme, are collapsed into single nodes resulting in an RG that is smaller and less complex. RGs were first developed for application in generic chemical structure retrieval using structure and substructure searching, but have since been applied to similarity searching [34-37].

A number of related methods have been published recently that report approaches to RGs for scaffold hopping. The general approach to reduced graph generation, as presented by Barker *et al.* [37], is given in Fig. (**9**).

Feature Trees

The Feature Trees (FTrees) approach [38]-[40] is very similar to the RG method, but encodes the nodes of the graphs based on their calculated physicochemical properties and, as the name suggests, the algorithm generally abstracts a given molecular graph to a tree; *i.e.* with no cyclic systems. Therefore, all rings are collapsed into single nodes, while fused or spiro ring systems may also be encoded as single nodes depending on the level of resolution used for the structure encoding. Steric features are also encoded to provide an indication of bulk properties of the molecule being encoded. The comparison of two trees is performed by a graph-matching algorithm; however, since the data structures are trees the comparisons are much more computationally tractable than tends to be the case with cyclic graphs. The application of FTrees to scaffold hopping was presented in [39] and was judged on a subjective basis to be capable of scaffold hopping by visual inspection.

Sheffield Reduced Graphs

Barker *et al.* [41] recently extended the extant work from Sheffield on RGs to scaffold hopping using a clique detection algorithm as a graph-matching comparison function; graph-matching on RGs was first reported by Takahashi *et al.* [34]. The graph reduction strategy applied in this study is similar to that published in [37] and given the designation ring/feature/link reduction/level 4 while also combining the identification of nodes that are acids or bases giving a total of 14 (17 when counting donor and acceptor nodes separately) RG node types. The graph-matching procedure is the

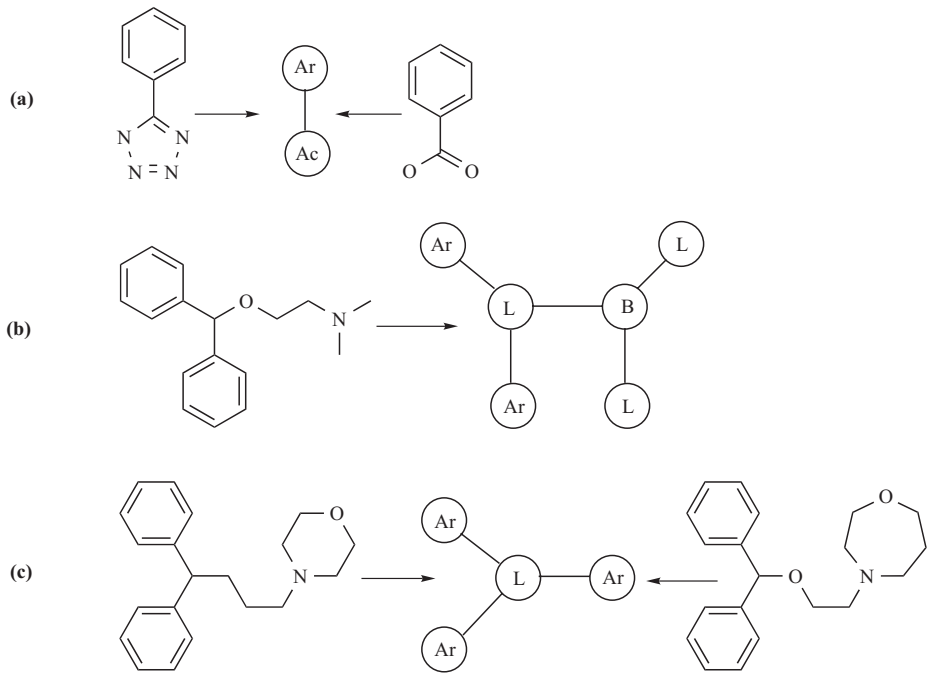


Fig. (9). Examples of reduced graphs: (a) acids take precedence over rings; (b) terminal nodes are described as linkers; (c) linker nodes may include heteroatoms if they have no hydrogen-bonding character. Ar: aromatic ring, no hydrogen-bonding; Ac: acid feature; L: acyclic inert node; B: base feature; RA: alicyclic ring, hydrogen bond acceptor. The figure has been redrawn from Barker *et al.* [37].

Bron-Kerbosch clique-detection algorithm [42] to discover the Maximum Common Induced Subgraph (MCIS), that is the maximum common node-induced subgraph between the two graphs under consideration. This study discovered that using a graph-matching comparator function the results for scaffold hopping were superior to results when using a fingerprint representation of the RGs and was better in many of the activity classes than the Daylight hash-key fingerprints.

Extended Reduced Graphs

The Extended Reduced Graphs (ErG) from Stiefl *et al.* [19], as the name suggests, extends the existing RG definition from the studies conducted at Sheffield. The extensions take the form of a combination of the RGs together with binding property pairs proposed by Kearsley *et al.* [43]. The first steps of the algorithm assign charges to the atoms of the molecule, followed by defining hydrogen-bonding. The atoms in the molecule are then abstracted as nodes that define their potential binding properties. The later stages of the algorithm then proceed to abstract the ring systems in the graph as the ring centroids for the general ring property features – *e.g.* aromatic – while retaining all bridgehead nodes and connecting these bridgeheads to their respective ring centroids. This approach also introduces a method of abstracting distance constraints to avoid issues in potential brittleness when considering identical node pairs with different edge distances between them that would otherwise be considered as dissimilar with many existing methods. A number of additional special cases are defined by Stiefl *et al.* and the reader is referred to [19] for the full definition of the algorithm.

4.4. Topomers

The topomer concept was introduced by Cramer *et al.* [44]-[48] for the rapid searching of large virtual combinatorial libraries *in silico*. Essentially, topomers are topologies of substructures from which a theoretically valid geometry can be generated *in silico* using a 3D generation program such as Concord [49]. The pre-defined attachment point of the substructure is then aligned with a vector in Cartesian space with consideration given to the chirality and torsion angle between the fragments. Usually only a single conformer of each of the fragments is used. The comparison of topomers is then achieved by examining the steric field properties of the topomers. Therefore, the similarity of the individual comparator topomers gives the overall similarity of the virtual molecules under consideration. The topomer approach was demonstrated to be effective for scaffold (or lead) hopping in a number of cases considering the low Tanimoto similarities of their fingerprints and also by visual inspection. Tripos, Inc. applies the topomer concept in analyzing a vast virtual library of approximately 10^{13} molecules as part of the ChemSpace™ platform.

4.5. FBSS

Field-Based Similarity Searching (FBSS) is a program that was developed in Sheffield to compare molecules using a Genetic Algorithm (GA) to optimize the alignments of the molecules' electrostatic potentials [50], [51], although alternative molecular fields can be used such as steric or hydrophobic fields. The GA optimizes the alignment of two molecules based on chromosomes representations that encode the translation and rotation of a given molecule against a molecule that is held in position. An individual candidate solution is scored according to the Gaussians calculated by the Carbó coefficient of the overlap of the molecular fields [52]. An example of an FBSS alignment is provided in Fig. (**10**). Schuffenhauer *et al.* demonstrated the potential of FBSS for scaffold hopping in reference [53]. Bohl *et al.* have also reported the application of FBSS for searching for novel yet equivalent scaffolds ensuring that the scaffold attachment points using steric fields as the comparator [4].

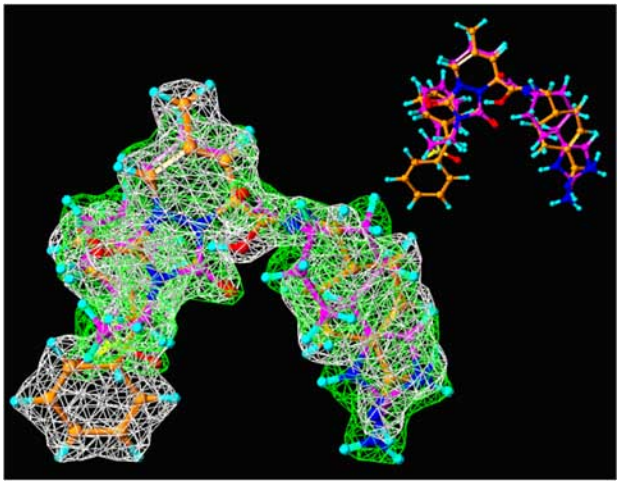


Fig. (10). Here are two thrombin inhibitor molecules with the steric fields used to overlay them in FBSS. The steric field of the target molecule, 1C4V (orange framework), is shown in white and the steric field of the hit molecule, 1D6W (magenta framework), is shown in green, both at 2 kcal/mol. The similarity determined by the Carbó coefficient was 0.832. The molecules are coded using the PDB (Protein Data Bank) system. Image courtesy of Kirstin Moffat of the University of Sheffield and reprinted here with permission.

4.6. Extended Electron Distribution Force Fields

Cresset Biomolecular Discovery [54] has developed eXtended Electron Distribution (XED) force fields determined based on the interaction potential of a probe molecule [55], [56]. The field point generation consists of a number of stages to characterize four different kinds of field properties: electrophilicity, electrophobicity (or nucleophilicity), van der Waals attractive (referred to as 'sticky' points), and hydrophobicity. The steric and electrostatic fields are first generated using a probe molecule to better define the interaction points available to other molecules. One hundred and twenty points are generated for each atom and these are used to define common extrema values to determine the final interaction points. Conformational flexibility is also taken into account by taking controlled subsets of the conformer space up to a defined maximum subset size.

The search process initially a distance matrix is generated of all of the pairwise Euclidean distances in ångströms between the field points allowing for direct comparisons to be

made between the field point sets of different molecules as a fast screening strategy. Only the top ranked comparisons are then considered for the more computationally-intensive field point alignment method using clique detection and simplex optimization algorithms. The Cresset approach has been demonstrated to be applicable in scaffold hopping [57] to novel chemotypes by visual inspection of the molecules retrieved, as illustrated in Fig. (**11**).

4.7. OpenEye Scientific Software

OpenEye has released two software programs for comparison of small molecules and can both be applied to *in silico* scaffold hopping campaigns. The first is ROCS (Rapid Overlay of Chemical Structures) and considers volume overlap as a measure of similarity, while the second is EON and considers the similarity of the electrostatic fields [58]. The programs can be used as part of a defined workflow, with ROCS providing an optimized alignment based on volume overlap, and EON calculating the electrostatic Tanimoto similarity of the two aligned structures.

ROCS

The ROCS algorithm attempts to discover the largest 3D volume overlap between two structures under consideration [59]. Like FBSS, ROCS also uses a Gaussian representation of the volume as opposed to the hard-sphere volume approach, but optimizes on the entire volume rather than the surface model. Since the product of Gaussians is also a Gaussian, the intersections are trivial to calculate. The comparisons are performed by application of the shape-based variants of the Tanimoto and Tversky similarity measures. The Tversky coefficient is asymmetric and particularly effective when considering molecules that are substantially different in size.

The speed of the ROCS approach typically permits 600-800 comparisons to be made per second. Therefore, multiple conformers of individual molecules can be considered as the normal method of analysis.

EON

The EON program calculates the similarity of two pre-aligned molecules based on the electrostatic fields of the two molecules [60]. This provides for a realistic comparison than simply based on shape, since it provides an objective measurement of the potential interaction similarities of the two structures under consideration.

4.8. FEPOPS

The calculation of the FEature POint PharmacophoreS (FEPOPS) descriptors proceeds by systematically generating a number of potential molecular geometries from the ligand alone [61]. Each conformer is then reduced to a 4-point pharmacophore using a K-medoids clustering algorithm, reducing *N* atoms to 4 feature points. The feature points are then assigned properties based on the interpolations of the atomic properties.

Although this method is considered for scaffold hopping, the approach does not explicitly consider a scaffold definition. Rather, it takes the somewhat inverse approach in that it attempts to disconnect the underlying scaffold of the molecule by looking at molecular graph vectors between atoms of functional interest in geometric space.

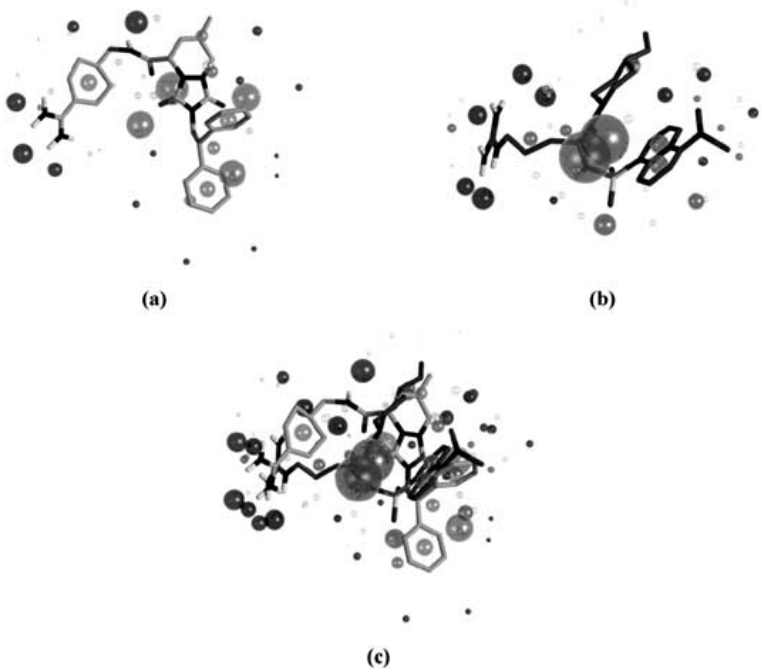


Fig. (11). Here two thrombin inhibitor molecules - (a) 1C4V and (b) 1FPC - with significantly different scaffolds are aligned (c) based on their field points, as determined by the XED approach. The field points correspond to electrophilicity, nucleophilicity, van der Waals attractive, and hydrophobicity. The sizes of the field point glyphs indicate the depth of the energy well for each point. Image courtesy of Cresset Biomolecular Discovery, Ltd. and reprinted here with permission.

4.9. MOLPRINT 3D

The MOLPRINT 3D descriptor is a geometric extension to the MOLPRINT 2D descriptor [62] by Bender *et al.* [63]. The algorithm proceeds by determining points on the surface of a given single conformation of a molecule. Interaction energies are then calculated for each of these surface points using theoretical probes that vary according to the particular interaction type under consideration. The probe types used are those defined by the GRID software [64]: C3, DRY, N1+, N2, O, and O-.

From each of these surface points, it is then possible to enumerate an environment of additional surface points radiating from the center point. A binary vector is defined for each radius from a given surface point. The EUs of the surface points are encoded into binary vectors for each of the probe types that encodes the EU values as bins with a bit being set for the surface point at that radius according to the binning scheme for that EU value as in Fig. (**12**).

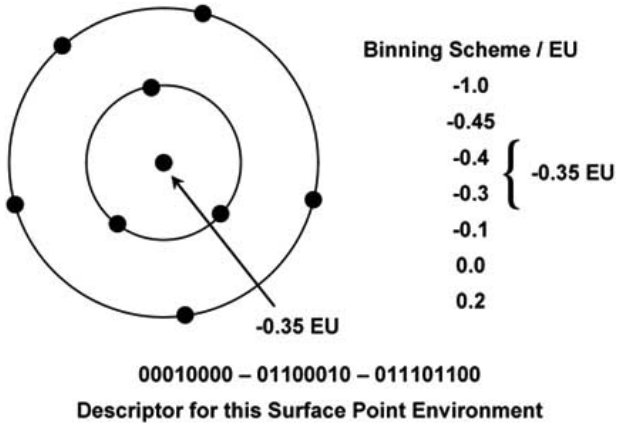


Fig. (12). Illustration of the descriptor generation step for a single surface point environment. The three components of the descriptor characterize the 0^{th}, 1^{st}, and 2^{nd} surface point environment layers, respectively. The figure has been redrawn from Bender *et al.* [36].

5. DISCUSSION AND CONCLUSIONS

Each of the methods described herein attempts to reduce the influence of the underlying connectivity of the molecules in the descriptor representation. This level of abstraction does not necessarily extend to the distances between those atoms, which tend to be defined rigidly based on the shortest bond distance between two atoms, although some of the most recent work does consider this [19]. Rigid adherence to this through-graph distance means that keys that are identical in all but distance, will be treated differently although there are many examples in which an increase or decrease in the distance will also fit the desired conformation and potential interactions. Therefore, it seems that the topological distances between atoms should be considered in all approaches as a further component to introduce a degree of obfuscation to reduce this rigorous adherence.

From the selection of 3D methods presented here, it seems that careful consideration of the conformational space is essential to enhance ligand-based scaffold hopping results. It is intuitive that the set of conformers selected should be representative of the entire conformational space; however it is not currently clear as to which method to apply in generating such a set of conformers. Alignment-free descriptor vectors [23, 65-67] appear to be the pre-eminent choice for the conformer problem since extant 'diverse' subset selection methods can be applied readily to select conformers that are representative of the entire conformer space and will be effective both within and between individual molecules.

The most recent advances in ligand-based scaffold hopping [24] using topology, geometry, and surface models have suggested that it is beneficial to apply a set of these complementary methods to ensure that the disjoint nature of these approaches allow for a more diverse set of returned structures in terms of their perceived chemotypes. While the sets of returned structures may be applied individually, there is a definite interest in combining results from such disparate representation methods using the data fusion approach – also referred to as consensus scoring in the virtual docking community – to generate more robust enrichments of retrieved molecules from diverse chemotypes [68].

In Chemoinformatics it is common to apply the graphical abstractions of molecules when considering automated analyses of large datasets to investigate the chemical class memberships of these molecules. However, there is no valid claim that these methods of molecular abstraction identify correctly the most salient points of the molecules under consideration in terms of their perceived scaffolds. Indeed, studies in which computational methods are applied in the field of scaffold hopping tend to claim scaffold hopping when some arbitrary limit is reached in dissimilarity between the molecules. This is by no means a criticism of these studies, merely a reflection of the current state of our scaffold definitions.

The major issue, as we see it, with scaffold definitions is the subjectivity involved in defining a scaffold. As Chemoinformaticians we prefer an invariant rule-based method that defines accurately and consistently the scaffold of a given molecule in isolation thereby permitting global application. However, the molecular scaffold is not simply a function of the molecule itself, since this information alone does not clarify what is important in terms of binding to a specific site on a protein.

A scaffold is not only dependent on the molecule itself, but it is also necessary to place that molecule in the context of the dataset of interest. For instance, blind application of the Bemis and Murcko scaffold determination algorithm, can often result in scaffolds that are uninformative or even misleading in instances with acyclic structures and cyclic substituents. However, the Bemis and Murcko frameworks do provide value in our chemical class analyses at present and recent research extends these concepts to increase further their application. Indeed, the Bemis and Murcko provide us with perhaps the most objective measure of success in scaffold hopping campaigns. Of course this does not imply that these methods should necessarily be applied in isolation, on the contrary the application of a number of approaches is advised.

One of the most important activities in Chemoinformatics is to generate reliable descriptions of molecules that contain

as much of the information related to that molecule as is possible. However, in the context of scaffolds and scaffold hopping, it is preferable that the descriptors are not as information-rich since this will tend to not find novel chemotypes in which we are interested. Therefore, it is important to develop descriptor generation methods that characterize the properties of the molecule in question, but not consider so strongly the scaffold component of the molecule. It is evident, therefore, that the preferred descriptors for scaffold hopping are those that introduce an element of fuzziness into the characterization.

The descriptors reviewed in this paper each provide abstracted descriptions of molecules in which the underlying connectivity of the chemical structure is either generalized or ignored completely. This process relaxes our definition of the neighborhood property in our chemical space of interest. In similarity searching it is the aim to locate molecules in the connectivity chemistry space with anticipation that these molecules will share similar properties. However, the aim in designing descriptors for scaffold hopping is to alter our definition of chemistry space so as to locate molecules that are similar to our query but not in terms of underlying connectivity.

The topological methods reviewed here tend to abstract the atoms of the molecular graph or indeed groups of atoms and look at their potential geometric distances based on their through-graph shortest distances. The geometric methods either consider physical properties of the atoms or surface aspects. However, very few of these methods consider the conformer problem and this seems essential in dealing with geometry. By considering only a single conformer generated *in silico* with no information *a priori* then the importance of that particular 3D model may be brought into doubt. Indeed, a subtle change in the 3D model of a ligand can lead to vastly different properties such as the Connolly surface. Therefore, it is anticipated that ongoing research will change the focus of these methods onto dealing with conformational flexibility and ensembles of conformers with special consideration for the extent to which these affect the potential for scaffold hopping.

It is our conclusion that current scaffold determination algorithms are prone to limitations that may eventually be overcome, but these methods will nonetheless rely on subjective interpretation. However, the extant approaches such as the frameworks from Bemis and Murcko do provide the practitioner with pragmatic approaches to classify molecules based on their chemotypes and further methods will undoubtedly be developed that will improve upon these approaches. The limitations to the existing methods detailed here require careful consideration and new methods are emerging that address these. Furthermore, new methods will continue to be developed to better consider molecular scaffolds both in their determination and the objective analysis of scaffold hopping campaigns to resolve more accurately whether the methods result in a scaffold hop.

ACKNOWLEDGEMENTS

The authors would like to thank Drs. Kamal Azzaoui, Peter Ertl, Richard A. Lewis, Stephen P. Jelfs, Jeremy L. Jenkins, Christian N. Parker, Ansgar Schuffenhauer, all Novartis Institutes for BioMedical Research. We would also like to thank Prof. Peter Willett and Dr. Val Gillet from the University of Sheffield, UK and Prof. Karl-Heinz Altmann of the ETH-Zürich, Switzerland.

REFERENCES

[1] Schuffenhauer, A.; Popov, M.; Schopfer, U.; Acklin, P.; Stanek, J.; Jacoby, E. *Comb. Chem. High Throughput Screen.* **2004**, *7*, 771.
[2] Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. *Angew. Chem., Int. Ed.* **1999**, *38*, 2894.
[3] Stanton, D. T.; Morris, T. W.; Roychoudhury, S.; Parker, C. N. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 21.
[4] Bohl, M.; Dunbar, J.; Gifford, E. M.; Heritage, T.; Wild, D. J.; Willett, P.; Wilton, D. J. *Quant. Struct. Act. Relat.* **2002**, *21*, 590.
[5] Böhm, H.-J.; Flohr, A.; Stahl, M. Scaffold hopping. *Drug Discov. Today Technologies* **2004**, *1*, 217.
[6] Markush, E. A. *Pyrazolone Dye and Process of Making the Same*. U.S. Patent No. 1,506,316, Granted: 26th August, **1924**.
[7] Reich, H. J.; Cram, D. J. *J. Am. Chem. Soc.* **1969**, *91*, 3527.
[8] Bemis, G. W.; Murcko, M. A. *J. Med. Chem.* **1996**, *39*, 2887.
[9] Roberts, G.; Myatt, G. J.; Johnson, W. P.; Cross, K. P.; Blower, Jr., P. E. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1302.
[10] Nicolau, C. A.; Tamura, S. Y.; Kelley, B. P.; Bassett, S. I.; Nutt, R. F. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1069.
[11] Cases, M.; Garciá-Serna, R.; Hettne, K.; Weeber, M.; van der Lei, J.; Boyer, S.; Mestres, J. *Curr. Top. Med. Chem.* **2005**, *5*, 763.
[12] Koch, M. A.; Schuffenhauer, A.; Scheck, M.; Wetzel, S.; Casaulta, M.; Odermatt, A.; Ertl, P.; Waldmann, H. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 17272.
[13] Lloyd, D. G.; Buenemann, C. L.; Todorov, N. P.; Manallack, D. T.; Dean, P. M. *J. Med. Chem.* **2004**, *47*, 493.
[14] Schneider, G.; Fechner, U. *Nat. Rev. Drug Discov.* **2005**, *4*, 649.
[15] Schneider, G.; Böhm, H.-J. *Drug Discov. Today* **2002**, *7*, 64.
[16] Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. *PROTEINS: Structure, Function, and Genetics* **2002**, *47*, 409.
[17] Ferrara, P.; Curioni, A.; Vangrevelinghe, E.; Meyer, T.; Mordasini, T.; Andreonin, W.; Acklin, P.; Jacoby, E. *J. Chem. Inf. Mod.* **2006**, *46*, 254.
[18] Good, A. C.; Cheney, D. L.; Sitkoff, D. F.; Tokarski, J. S.; Stouch, T. R.; Bassolino, D. A.; Krystek, S. R.; Li, Y.; Mason, J. S.; Perkins, T. D. J. *J. Mol. Graph. Mod.* **2003**, *22*, 31.
[19] Stiefl, N.; Watson, I. A.; Baumann, K.; Zaliani, A. *J. Chem. Inf. Mod.* **2006**, 46, 208.
[20] Xu, Y.-J.; Johnson, M. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 912.
[21] Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A. *J. Chem. Inf. Mod.* **2006**, *46*, 462.
[22] Godden, J. W.; Stahura, F. L.; Bajorath, J. *J. Chem. Inf. Mod.* **2005**, *45*, 1812.
[23] Renner, S.; Schneider, G. *Chem. Med. Chem.* **2006**, *1*, 181.
[24] Broto, P.; Moreau, G.; Vandyke, C. *Eur. J. Med. Chem.* **1984**, *19*, 66.
[25] Fechner, U.; Franke, L.; Renner, S.; Schneider, P.; Schneider, G. *J. Comput. Aided Mol. Des.* **2003**, *17*, 687.
[26] CORINA is available from Molecular Networks, GmbH at http://www.mol-net.de.
[27] Bush, B. L.; Sheridan, R. P. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 756.
[28] The Molecular Operating Environment (MOE) is available from Chemical Computing Group, at http://www.chemcomp.com/.
[29] Renner, S.; Ludwig, V.; Boden, O.; Scheffer, U.; Göbel, M.; Schneider, G. *Chembiochem,* **2005**, *6*, 1119.
[30] Renner, S.; Schneider, G. *J. Med. Chem.* **2004**, *47*, 4653.
[31] Schuffenhauer, A.; Floersheim, P.; Acklin, P.; Jacoby, E. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 391.
[32] Gillet, V. J.; Downs, G. M.; Ling, A.; Lynch, M. F.; Venkataram, P.; Wood, J. V.; Dethlefsen, W. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 126.
[33] Gillet, V. J.; Downs, G. M.; Holliday, J. D.; Lynch, M. F.; Dethlefsen, W. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 260.
[34] Takahashi, Y.; Sukekawa, M.; Sasaki, S. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 639.
[35] Fisanick, W.; Lipkus, A. H.; Rusinko, A., III. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 130.

[36] Gillet, V. J.; Willett, P.; Bradshaw, J. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 338.
[37] Barker, E. J.; Gardiner, E. J.; Gillet, V. J.; Kitts, P.; Morris, J. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 346.
[38] Rarey, M.; Dixon, J. S. *J. Comput. Aided Mol. Des.* **1998**, *12*, 471.
[39] Rarey, M.; Stahl, M. *J. Comput. Aided Mol. Des.* **2001**, *15*, 497.
[40] Lengauer, T.; Lemmen, C.; Rarey, M.; Zimmermann, M. *Drug Discov. Today* **2004**, *9*, 27.
[41] Barker, E. J.; Cosgrove, D. A.; Gardiner, E. J.; Gillet, V. J.; Kitts, P.; Willett, P. *J. Chem. Inf. Mod.* **2006**, *46*, 503.
[42] Bron, C.; Kerbosch, J. *Commun. ACM* **1973**, *16*, 575.
[43] Kearsley, S. K.; Sallamack, S.; Fluder, E. M.; Andose, J. D.; Mosley, R. T.; Sheridan, R. P. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 118.
[44] Cramer, R. D.; Patterson, D. E.; Clark, R. D.; Soltanshahi, F.; Lawless, M. S. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 1010.
[45] Cramer, R. D.; Poss, M. A.; Hermsmeier, M. A.; Caulfield, T. J.; Kowala, M. C.; Valentine, M. T. *J. Med. Chem.* **1999**, *42*, 3919.
[46] Andrews, K. M.; Cramer, R. D. *J. Med. Chem.* **2000**, *43*, 1723.
[47] Cramer, R. D.; Jilek, R. J.; Andrews, K. M. *J. Mol. Graph. Model.* **2002**, *20*, 447.
[48] Cramer, R. D.; Jilek, R. J.; Guessregen, S.; Clark, S. J.; Wendt, B.; Clark, R. D. *J. Med. Chem.* **2004**, *47*, 6777.
[49] Concord is available from Tripos, Inc. at http://www.tripos.com.
[50] Wild, D. J.; Willett, P. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 159.
[51] Thorner, D. A.; Wild, D. J.; Willett, P.; Wright, P. M. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 900.
[52] Good, A. C.; Richards, W. G. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 112.
[53] Schuffenhauer, A.; Gillet, V. J.; Willett, P. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 295.

Received: April 03, 2006 **Revised: April 24, 2006** **Accepted: April 25, 2006**

[54] The FieldScreen, FieldTemplater, FieldAlign, and XedeX software are available from Cresset BioMolecular Discovery Ltd. at http://www.cresset-bmd.com/.
[55] Kalindjian, S. B.; Buck, I. M.; Davies, J. M. R.; Dunstone, D. J.; Hudson, M. L.; Low, C. M. R.; McDonald, I. M.; Pether, M. J.; Steel, K. I. M.; Tozer, M. J.; Vinter, J. G. *J. Med. Chem.* **1996**, *39*, 1806.
[56] Cheeseright, T.; Mackey, M.; Rose, S.; Vinter, A. *J. Chem. Inf. Mod.* **2006**, *46*, 665.
[57] Low, C. M. R.; Buck, I. M.; Cooke, T.; Cushnir, J. R.; Kalindjian, S. B.; Kotecha, A.; Pether, M. J.; Shankley, N. P.; Vinter, J. G.; Wright, L. *J. Med. Chem.* **2005**, *48*, 6790.
[58] ROCS and EON are both available from OpenEye Scientific Software, Inc. at http://www.eyesopen.com.
[59] Rush, T. S., III.; Grant, J. A.; Mosyak, L.; Nicholls, A. *J. Med. Chem.* **2005**, *48*, 1489.
[60] Nicholls, A.; MacCuish, N. E.; MacCuish, J. D. *J. Comput. Aided Mol. Des.* **2004**, *18*, 451.
[61] Jenkins, J. L.; Glick, M.; Davies, J. W. *J. Med. Chem.* **2004**, *47*, 6144.
[62] Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 170.
[63] Bender, A.; Mussa, H. Y.; Gill, G. S.; Glen, R. C. *J. Med. Chem.* **2004**, *47*, 6569.
[64] Goodford, P. J. *J. Med. Chem.* **1985**, *28*, 849.
[65] Hemmer, M. C.; Steinhauer, V.; Gasteiger, J. *Vib. Spectrosc.* **1999**, *19*, 151.
[66] Brown, N.; McKay, B.; Gasteiger, J. *QSAR Comb. Sci.* **2005**, *24*, 480.
[67] Hirons, L.; Holliday, J. D.; Jelfs, S. P.; Willett, P.; Gedeck, P. *QSAR Comb. Sci.* **2005**, *24*, 611.
[68] Willett, P. *J. Med. Chem.* **2005**, *48*, 4183.