Computational and Structural Biotechnology Journal 15 (2017) 86–90

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Mini Review

Matched Molecular Pair Analysis in Short: Algorithms, Applications and Limitations

Christian Tyrchan, Emma Evertsson *

Department of Medicinal Chemistry, RIA iMed, AstraZeneca R&D, Pepparedsleden 1, S-431 83 Mölndal, Sweden

ARTICLE INFO

Article history:
Received 28 October 2016
Received in revised form 8 December 2016
Accepted 9 December 2016
Available online 13 December 2016

ABSTRACT

Molecular matched pair (MMP) analysis has been used for more than 40 years within molecular design and is still an important tool to analyse potency data and other compound properties. The methods used to find matched pairs range from manual inspection, through supervised methods to unsupervised methods, which are able to find previously unknown molecular pairs. Recent publications demonstrate the value of automatic MMP analysis of publicly available bioactivity databases. The MMP concept has its limitations, but because of its easy to use and intuitive nature, it will remain one of the most important tools in the toolbox of many drug designers.

Contents

1. Introduction

The challenge of molecular design is the decision what to do next based on available data, medicinal chemistry expert knowledge, experience and intuition [1]. In small sets of molecules an experienced chemist can spot trends and relationships by eye. As the numbers of compounds increases, more systematical approaches are needed. Already in the early 70s, methods for systematic analysis were published e.g. the Topliss Scheme [1] or the Craig Plot [2], recommending a systematic stepwise method of building a structure–activity relationship for a chemical series. Hansch [3], Free and Wilson [4] reasoned in the 1960s that the biological activity for a set of analogues could be described by the contributions that substituents or structural elements make to the activity of a parent structure. Generally speaking these local QSAR methods try to find a correlation between structural and physicochemical descriptors towards a given endpoint [4], such as biological activity.

The term Molecular Matched Pair (MMP) was coined in 2004 by Kenny and Sadowski [5], for a special case of QSAR; now a widely used concept throughout drug design processes. In the most common situation, MMP describes a pair of compounds that differ structurally at a single site through a well-defined transformation (see Fig. 1) that is associated with a relative change in a property value. The correlation between the structural change and the property change is used in rationalizing observed structure–property-relationships (SPR) and compound optimization. Several different applications for MMP analysis originating from industry or academia have been developed and published, highlighting its importance. Among others these include: Drug-Guru [6,7], Buy me Grease [7], WizePairZ [8], T-Analyse and T-Morph [9], VAMMPIRE [10] as well as the Hussain-Rea MMP algorithm [11] (Table 1). The MMP concept has been further developed into Matched Pair Series [12,13] or Matched Molecular Series (MMS) [14] to describe a set of compounds (not only a pair) differing by only a single chemical transformation.

Recently an extension of the MMP concept towards biopharmaceutical applications was published, using macromolecular sequence data

* Corresponding author.
E-mail address: emma.evertsson@astrazeneca.com (E. Evertsson).

http://dx.doi.org/10.1016/j.csbj.2016.12.003

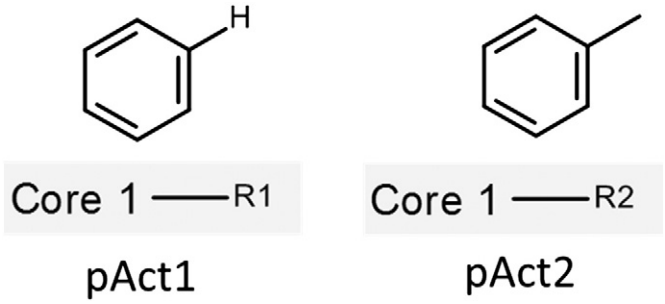

Fig. 1. Example of a matched molecular pair (MMP).

to predict the effect of single amino acid substitutions on property optimization [15].

Besides supporting hypothesis development and testing, an important application of MMP is in the detection of outliers, namely a pair of compounds that show a step change in a property; a so called activity cliff. These compounds are often the most interesting to study in the design of compounds targeting improvement of the property showing this change [16,17]. An inherently difficult problem to detect these activity cliffs is confounded by experimental uncertainties in the measured properties, since they are a function of the chemical space representation [18]. One systematic approach to the detection of activity cliffs and determination of their depth uses support vector regression [19]. Not only can different chemical space representations lead to significant changes in the nature of these activity cliffs, but even simple atomic variations can cause dramatic effects on important complex endpoints in medicinal chemistry; dose to man prediction, potency, clearance, solubility and permeability to name a few [18,20]. If the structural change (R group) is small and the scaffold in a chemical series is conserved, the MMP represents a relevant and easy to interpret chemical space representation. The MMP approach can further be extended to systematically analyse non-additivity in a structure property relationship (SPR) series [21].

2. Application and Limitation

The assumption that the effect of chemical substitution can be generalized, is inherently assumed in all QSAR methods, including the MMP approach, successfully highlighted by the work of Lipinski et al. who correlated physicochemical properties to oral bioavailability [22]. With the increasing availability of public databases containing millions of structure–activity-relationship (SAR) [23,24] or SPR data, multiple papers have been published applying MMP concept to: ADME [25,26], bioisosterism [9,27,28], aqueous solubility [29–32], plasma protein binding [29,30], oral exposure [29], logD [8,30,32], potency [8,9,27,31,33], intrinsic clearance [7,34], herG and P450 metabolism [29,32,34], in vitro UGT (Uridine 5′-diphospho-glucuronosyltransferase) glucuronidation clearance [35], half-life [31], selectivity against off-targets [31], impact of N- and O-methylation on aqueous solubility and lipophilicity [36] or mode of action; [31] the analysis differing only in the MMP algorithm used.

In two relatively recent publications [31,37] an apparently simple MMP transformation of CH → C-CH3 is analysed in greater detail and highlights some general limitations and drawbacks of using the MMP concept prospectively in drug design. The methyl group is a commonly occurring carbon fragment in small-molecule drugs and can modulate both the biological and physical properties of a molecule. Two literature analysis of >2000 cases of methylation revealed that an activity boost of a factor of 10 or more is found with an approximate 8% frequency, and the probability of achieving a 100-fold boost is less than 1% [33]. However, the distribution of potency changes in respect to the MMP is often nearly symmetrical and centred at or near zero resulting in a similar likelihood of causing potency gains or losses. A consistent bias of specific substituents towards improved potency could not be observed. Nevertheless an understanding of these rare events affecting the binding potency by improving the IC_{50} value of a compound by more than

Table 1
Classified MMP algorithms.

Non-supervised methods			
	R. Guha (2012) [46]	BCI structural fingerprints, CDK 1024-bit path fingerprint	2016
	Fuchs et al. (2015) [15]	Sequence alignment for peptide MMP[a]	2015
	T.J. Ritchie et al. (2015) [36]	HRF[b]	2015
	Matsy (2014) [13]	HRF[b]	2014
	VAMMPIRE (2013, 2014) [10,49]	MCS and HRF[b]	2013
	C.E. Keefer et al. (2011) [25]	Modified HRF[b] (Pairfinder)	2011
	J. Bajorath et al. (2010–2016) [12,14,19,27,43,50–52]	HRF[b], modified HRF[b], RECAP[c] fragmentation	2010
	J. Hussain et al. (2010) [11]	Hussain and Rea fragmentation (HRF[b])	2010
	L. Cururull-Sanchez (2010) [35]	ECFP6 fingerprints with sub-structure search	2010
	Papadatos et al. (2010) [42]	dt_commonsubstruct and findsub routine from Daylight and HRF[b]	2010
	WizePairs (2010) [8]	MCS[d] and SMIRKS[e]	2010
	Raymond et al. (2009) [47]	MSM[f] rule framework based on MCS[d]	2009
	R. Sheridan et al. (2002, 2006) [9,28]	Similarity and MCS[d] method (T-Analyse)	2008
Supervised methods	ThricePairs (2010) [34]	Defined transformations, SMARTS[g]	
	Gleeson et al. (2009) [26]	Substructure Search	2010
	Buy me Grease (2009, 2010) [7,35]	Defined transformations, RXN[h] format	2009
	P.J. Hajduk et al. (2008) [33]	Findsub routine from Daylight and defined transformations, SMIRKS[e]	2009
	D.Y. Haubertin et al. (2007) [30]	RECAP[c] method	2008
	Drug Guru (2006) [6]	Defined transformations, SMIRKS[e]	2007
	N.T. Southall et al. (2006) [53]	Topological torsion similarity and MCS[d]	2006
	A. G. Leach et al. (2006) [29]	Defined transformations, SMARTS[g] (Leatherface)	2006
	T.J. Ritchie (2016) [48].	SMIRKS[e]	2006

[a] MMP: Molecular Matched Pair.
[b] HRF: Hussain and Rea fragmentation.
[c] RECAP: Retrosynthetic Combinatorial Analysis Procedure.
[d] MCS: Maximum common substructure.
[e] SMIRKS: SMILES reaction specification.
[f] MSM: molecular substructure modification.
[g] SMARTS: SMiles ARbitrary Target Specification.
[h] RXN: MDL Molfile Reaction Format.