



## Matched molecular pair analysis of activity and properties of glycogen phosphorylase inhibitors

Alan M. Birch, Peter W. Kenny \*, Iain Simpson, Paul R. O. Whittamore

AstraZeneca, Mereside, Alderley Park, Macclesfield, Cheshire SK10 4TG, UK

### ARTICLE INFO

#### Article history:

Received 3 October 2008

Revised 27 November 2008

Accepted 2 December 2008

Available online 6 December 2008

#### Keywords:

Glycogen phosphorylase

Enzyme inhibitor

Diabetes

Matched molecular pair analysis

MMPA

Molecular editor

Cheminformatics

Leatherface

QSAR

QSPR

Physicochemical properties

Solubility

Plasma protein binding

Tetrazole

### ABSTRACT

Matched molecular pair analysis has been used in design of inhibitors of glycogen phosphorylase.

© 2008 Elsevier Ltd. All rights reserved.

Quantitative structure–activity/property relationships (QSAR/QSPR) are used to predict biological activity (e.g.,  $IC_{50}$ ) or physicochemical properties (e.g., solubility). A common approach to building these models is to encode structures using molecular descriptors and use techniques such as multilinear regression<sup>1</sup> (MLR) or partial least squares<sup>1</sup> (PLS) to model properties of interest. Although PLS addresses the problem of correlated descriptors in MLR, the resulting models are not easily interpreted. Overfitting<sup>2</sup> is a concern in multivariate QSAR/QSPR studies. Validation procedures may present an optimistic view of model quality and scope when families of close structural analogs are present in the data. Models may be termed global or local according to their intended scope although some global models would be more accurately described as ensembles of local models.

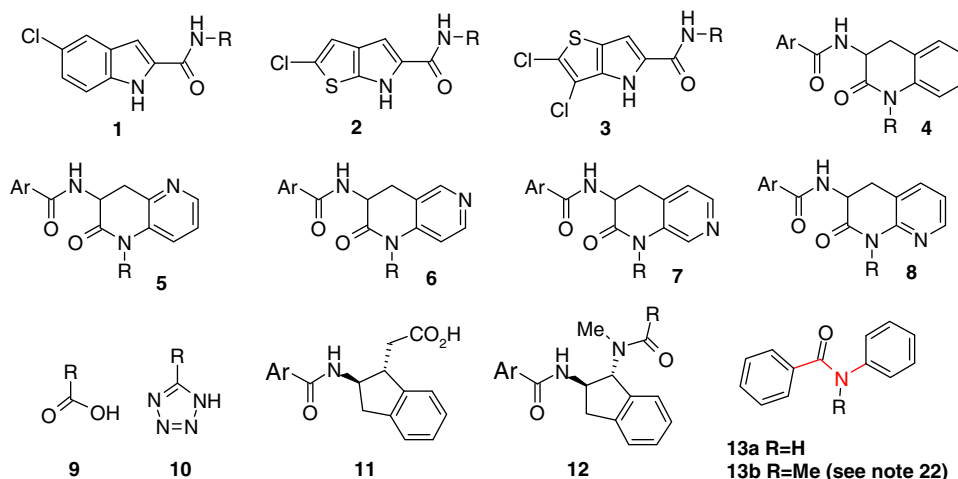
An alternative approach to deriving QSAR/QSPR models is to associate contributions to the property of interest with specific structural features. In Free–Wilson analysis<sup>3</sup> (FWA) the value of property of interest is partitioned into fragment contributions. Matched molecular pair analysis<sup>4</sup> (MMPA) focuses on the effects

of specific structural changes on properties of interest. The key assumption in MMPA is that differences in a property can be predicted more accurately than the property itself and an analogy can be drawn with the free energy perturbation method.<sup>5</sup> This distinguishes it from FWA to which it is superficially similar. In contrast to FWA, MMPA can be applied readily to structural diverse data sets which increases confidence that observed effects of structural changes are globally relevant.

MMPA quantifies differences in a property that are linked to a specific structural relationship. The relationship may be simple (e.g., chloro compounds and their des-chloro analogs) or more abstract such as that which exists between benzamides and the corresponding anilides. Structural relationships between molecules can be established by applying molecular transformations and used to find consistent local QSAR.<sup>6</sup> The principal outputs of MMPA are the mean difference in the property and the associated standard deviation, which can also be used to derive a confidence interval for the mean. MMPA can be seen as a special type of local QSAR in which a specific structural difference connects two diverse groups of structures. The matched molecular pair paradigm has been used in analysis of physicochemical properties,<sup>7,8</sup> oral exposure,<sup>7</sup> and potency.<sup>9</sup> Using examples from a glycogen

\* Corresponding author. Tel.: +44 1625 514396.

E-mail address: [pwk.pub.2008@gmail.com](mailto:pwk.pub.2008@gmail.com) (P.W. Kenny).



Standard deviation (SD) reflects both strength of the matched molecular pair relationship and precision of the assay. The lowest

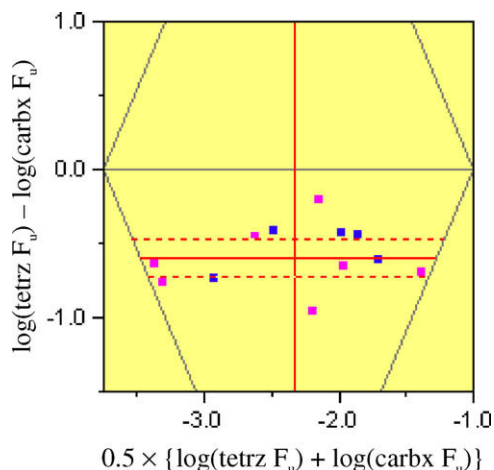
Another commonly used approach to increase aqueous solubility is to incorporate substituents that are ionized under normal physiological conditions. A number of carboxylic acids (e.g., **11**<sup>17</sup>), were synthesized and found to bind too strongly to plasma proteins. Bioisosteric replacements for the carboxylate were considered for reducing protein binding. MMPA was used to show that

**Table 2**  
Effects on aqueous solubility and potency of substituting nitrogen for aromatic carbon

Pair	Property	Mean <sup>a</sup>	SE <sup>b</sup>	SD <sup>c</sup>	N <sup>d</sup>	%Inc <sup>e</sup>
1 → 2	plC <sub>50</sub> <sup>f</sup>	0.21	0.06	0.46	57	68
1 → 3	plC <sub>50</sub> <sup>f</sup>	0.38	0.06	0.47	59	83
2 → 3	plC <sub>50</sub> <sup>f</sup>	0.29	0.07	0.48	54	78
1 → 2	logF <sub>u</sub> <sup>g</sup>	0.13	0.04	0.22	28	71
1 → 3	logF <sub>u</sub> <sup>g</sup>	-0.30	0.06	0.27	22	9
2 → 3	logF <sub>u</sub> <sup>g</sup>	-0.42	0.08	0.37	22	5
1 → 2	logS <sup>h</sup>	0.20	0.09	0.54	37	68
1 → 3	logS <sup>h</sup>	-0.29	0.13	0.61	22	45
2 → 3	logS <sup>h</sup>	-0.62	0.13	0.69	30	20

Pair	Property	Mean <sup>a</sup>	SE <sup>b</sup>	SD <sup>c</sup>	N <sup>d</sup>	%Inc <sup>e</sup>
CH → N <sup>f</sup>	log S <sup>g</sup>	0.54	0.09	0.92	108	72
<b>4</b> → <b>5</b>	log S <sup>g</sup>	1.00	0.21	0.47	5	100
<b>4</b> → <b>5</b>	plC <sub>50</sub> <sup>h</sup>	−0.48	0.14	0.43	10	10
<b>4</b> → <b>6</b>	log S <sup>g</sup>	0.55	0.23	0.47	4	75
<b>4</b> → <b>6</b>	plC <sub>50</sub> <sup>h</sup>	−1.60	0.41	1.01	6	0
<b>4</b> → <b>7</b>	log S <sup>g</sup>	0.32	0.24	0.53	5	80
<b>4</b> → <b>7</b>	plC <sub>50</sub> <sup>h</sup>	−0.77	0.12	0.39	10	0
<b>4</b> → <b>8</b>	log S <sup>g</sup>	0.52	0.28	0.62	5	80
<b>4</b> → <b>8</b>	plC <sub>50</sub> <sup>h</sup>	−1.21	0.13	0.33	7	0

<sup>a</sup> Mean difference in property.  
<sup>b</sup> Standard error in mean.  
<sup>c</sup> Standard deviation in mean.  
<sup>d</sup> Number of matched molecular pairs.  
<sup>e</sup> Percentage of pairs for which the value of the property increased.  
<sup>f</sup> Initial analysis; does not include aza-dihydroquinolones which were synthesized after this analysis had been performed.  
<sup>g</sup>  $\log_{10}(\text{solubility}/\text{mol dm}^{-3})$ ; solid stirred in 0.1  $\text{mol dm}^{-3}$  sodium phosphate buffer for 24 h at 25 °C and pH 7.4 and then centrifuged.  
<sup>h</sup>  $-\log_{10}(\text{IC}_{50}/\text{mol dm}^{-3})$ ; see Ref. 15.



**Figure 1.** Matched pair plot illustrating differences (mean =  $-0.60$ ; SE =  $0.06$ ; SD =  $0.20$ ) in fraction unbound ( $F_u$ ) for carboxylate–tetrazole pairs. Solid red lines show mean values for the quantities plotted and broken red lines define a 95% confidence interval for the difference. The seven matched pairs used in the initial analysis (mean =  $-0.64$ ; SE =  $0.09$ ; SD =  $0.23$ ) are colored in mauve and five matched pairs, for which data subsequently became available, are colored in blue.

**Table 3**  
Effects of N-methylation on aqueous solubility of secondary amides

Analysis	Pair	Mean <sup>a</sup>	SE <sup>b</sup>	SD <sup>c</sup>	N <sup>d</sup>	%Inc <sup>e</sup>
1a	Acyclic; aliphatic amine <sup>f</sup> ; original 2003 analysis	0.37	0.14	0.71	26	65
1b	Acyclic; aliphatic amine <sup>f</sup> ; all pairs	0.59	0.07	0.71	109	76
1c	Acyclic; aliphatic amine <sup>f</sup> ; indans	0.54	0.11	0.71	38	74
1d	Acyclic; aliphatic amine <sup>f</sup> ; no indans	0.61	0.08	0.72	71	82
2	Benzanilides <sup>g</sup>	1.49	0.15	0.47	9	100
3	Cyclic amides <sup>h</sup>	0.18	0.25	0.76	9	44

<sup>a</sup> Mean difference in  $\log_{10}(\text{solubility}/\text{mol dm}^{-3})$ .

<sup>b</sup> Standard error in mean.

<sup>c</sup> Standard deviation of mean.

<sup>d</sup> Number of pairs.

<sup>e</sup> Percentage of pairs for which increase in solubility is observed.

<sup>f</sup> SMARTS: [CH3]-N([CX4])-[C@H](O)[#6].

<sup>g</sup> SMARTS: [CH3]-[N;X3;1R](-c1cccc1)-[C;1R](=O)-c2cccc2.

<sup>h</sup> SMARTS: [CH3]-[NX3]([C&X4,c])-[C@H](O)[C&X4,c].

replacement with tetrazole, a commonly used carboxylate bioisostere<sup>18</sup> was likely to increase binding to plasma protein (Fig. 1). Consequently, no tetrazoles were synthesized in this series.

After only a few compounds had been synthesized in the indan series (**12**<sup>19</sup>), one of us (IS) observed that N-methylation of the amide appeared to improve aqueous solubility. This is counterintuitive because a hydrogen bond donor is removed. MMPA was used to investigate whether this behavior was general or specific to the indan series. The initial analysis (1a, Table 3), performed in mid-2003 using data for 26 matched pairs, showed that N-methylation of the amide led to a mean increase in solubility of 0.37 log units. While the initial analysis suggested that the observed effect of N-methylation was both real and not restricted to the indan series, there were concerns about the size of the sample and the diversity of the structures. The analysis was repeated in mid-2008 for this study using a larger data set that included measurements made specifically for a matched molecular pair study<sup>7</sup> (analysis 1b). No significant difference was observed between indans (analysis 1c) and the other compounds (analysis 1d) in this data set.

The contrasting effects of N-methylation on solubility of benzanilides and cyclic amides illustrate the need for precise specifica-

tion of substructural context in MMPA. N-Methylation of benzanilides leads to a mean increase in solubility of 1.49 log units (analysis 2) which is significantly greater than the corresponding figure for acyclic amides derived from aliphatic amines (analysis 1b). In contrast, N-methylation of cyclic amides led to an increase in solubility in less than half of the matched molecular pairs and the mean change in solubility does not differ significantly from zero.

The similarity of the 1-octanol/water partition coefficients<sup>20</sup> ( $\log P_{\text{oct}}$ ) of *N,N*-dimethylacetamide ( $-0.77$ ) and *N*-methylacetamide ( $-1.05$ ) might be invoked as evidence that the hydrogen bond donor of the latter contributes little to its aqueous solvation. However, 1-octanol forms hydrogen bonds with solutes and hexadecane/water partition coefficients ( $\log P_{\text{hxd}}$ ) are more relevant. Using  $\log P_{\text{oct}}$  (2.24) and  $\log P_{\text{hxd}}$  (0.66) measured<sup>21</sup> for indole as a guide suggests that methylation of *N*-methylacetamide is likely to lead to an increase in  $\log P_{\text{hxd}}$  of between 1.5 and 2.0. The origins of the effect on solubility appear to lie in the packing of molecules in the solid state.<sup>7</sup>

The hydrogen bond donor and acceptor sit on opposite sides of the amide group in acyclic secondary amides. This geometry that allows the amide groups to form infinite ladders of hydrogen bonds in the solid state.<sup>7</sup> The *cis* geometry typically adopted by small (less than 9-membered) cyclic amides does not allow infinite ladders of hydrogen bonds to be formed as easily so the donor contributes less to stability of the solid state of these species. The larger effect of N-methylation on solubility of benzanilides probably reflects increased stability<sup>22</sup> of the *cis* amide conformation in *N*-methylanilides.

The examples of MMPA in this study focus on average differences in properties. The analytic focus can easily be shifted to matched molecular pairs for which the largest differences in property are observed. These ‘activity cliffs’<sup>23,24</sup> could be used to determine limits for potency gains that can be achieved with specific structural changes. This information might be used to assess how efficiently<sup>25–27</sup> compounds exploit their molecular recognition elements in binding to targets.

A general and completely automated method for identification of matched molecular pairs has not yet been reported. In the work described in this and other studies,<sup>4,7–9</sup> it was necessary to define the structural relationship of each matched molecular pair at the start of the analysis or make assumptions about its nature (e.g., side chain interchange<sup>8</sup>). It may prove possible to generate useful representations of relationships between structures directly from the structures themselves. Such representations will need to be unique, like canonical SMILES,<sup>12</sup> in order to be useful and their generation presents a significant cheminformatic challenge.

## Acknowledgments

The challenging syntheses of compounds in the series **5**, **6**, **7**, and **8** were performed by Andy Stocker, Adrian Pickup, and Cyril Dousson. Solubility and protein binding were measured by members, past and present, of the Alderley Park Physical Chemistry Group.

## References and notes

- Kalivas, J. H. *J. Chemometrics* **1999**, *13*, 111.
- Hawkins, D. M. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1.
- Free, S. M.; Wilson, J. W. *J. Med. Chem.* **1964**, *7*, 395.
- Kenny, P. W.; Sadowski, J. In *Methods and Principles in Medicinal Chemistry*; Oprea, T. I., Ed.; Wiley-VCH: Weinheim, 2005; Vol. 23, pp 271–285.
- Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420.
- Sheridan, R. P.; Hunt, P.; Culberson, C. J. *J. Chem. Inf. Model.* **2006**, *46*, 180.
- Leach, A. G.; Jones, H. D.; Cosgrove, D. A.; Kenny, P. W.; Ruston, L.; MacFaul, P.; Wood, J. M.; Colclough, N.; Law, B. J. *Med. Chem.* **2006**, *49*, 6672.

8. Haubertin, D. Y.; Bruneau, P. J. *Chem. Inf. Model.* **2007**, 47, 1294.
9. Hajduk, P. J.; Sauer, D. R. *J. Med. Chem.* **2008**, 51, 553.
10. <http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html>.
11. Weininger, D. *J. Chem. Inf. Comput. Sci.* **1988**, 28, 31.
12. Weininger, D.; Weininger, A.; Weininger, J. L. *J. Chem. Inf. Comput. Sci.* **1989**, 29, 97.
13. Whittamore, P. R. O.; Addie, M. S.; Bennett, S. N. L.; Birch, A. M.; Butters, M.; Godfrey, L.; Kenny, P. W.; Morley, A. D.; Murray, P. M.; Oikonomakos, N. G.; Otterbein, L. R.; Pannifer, A. D.; Parker, J. S.; Readman, K.; Siedlecki, P. S.; Schofield, P.; Stocker, A.; Taylor, M. J.; Townsend, L. A.; Whalley, D. P.; Whitehouse, J. *Bioorg. Med. Chem. Lett.* **2006**, 16, 5567.
14. Birch, A. M.; Kenny, P. W.; Oikonomakos, N. G.; Otterbein, L.; Schofield, P.; Whittamore, P. R. O.; Whalley, D. P. *Bioorg. Med. Chem. Lett.* **2007**, 17, 394.
15. Freeman, S.; Bartlett, J. B.; Convey, G.; Hardern, I.; Teague, J. L.; Loxham, S. J. G.; Allen, J. M.; Poucher, S. M.; Charles, A. D. *Br. J. Pharmacol.* **2006**, 149, 775.
16. Perrin, D. D. *Dissociation Constants of Organic Bases*; Butterworths: London, 1965.
17. Birch, A. M.; Johnstone, C.; Plowright, A. T.; Simpson, I.; Whittamore, P. R. O. WO 082400, 2006.
18. Herr, R. J. *Bioorg. Med. Chem.* **2002**, 10, 3379.
19. Whittamore, P. R. O.; Bennett, S. N. L.; Simpson, I. WO 074484, 2003.
20. Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR Hydrophobic, Electronic, and Steric Constants*; American Chemical Society: Washington, DC, 1995.
21. Toulmin, A.; Wood, J. M.; Kenny, P. W. *J. Med. Chem.* **2008**, 51, 3720.
22. Conformational preferences of benzanilides were explored by searching the Cambridge Structural Database (<http://www.ccdc.cam.ac.uk/products/csd>) for the fragments **13a** and **13b** and analyzing distributions of the dihedral angle,  $\tau$ , defined by the bonds colored red. The value of  $\text{abs}(\tau)$  exceeded  $160^\circ$  for all 485 matches (292 refcodes) with **13a** but was less than  $30^\circ$  for most (54/66 fragments; 6/36 refcodes) matches with **13b**.
23. Maggiora, G. M. *J. Chem. Inf. Model.* **2006**, 46, 1535.
24. Guha, R.; Van Drie, J. H. *J. Chem. Inf. Model.* **2008**, 48, 646.
25. Andrews, P. R.; Craik, D. J.; Martin, J. L. *J. Med. Chem.* **1984**, 27, 1648.
26. Hopkins, A. L.; Groom, C. R.; Alex, A. *Drug Discovery Today* **2004**, 9, 430.
27. Verdonk, M. L.; Rees, D. C. *ChemMedChem* **2008**, 3, 1179.