Matched Molecular Pairs as a Guide in the Optimization of Pharmaceutical Properties; a Study of Aqueous Solubility, Plasma Protein Binding and Oral Exposure

Andrew G. Leach,* Huw D. Jones, David A. Cosgrove, Peter W. Kenny, Linette Ruston, Philip MacFaul, J. Matthew Wood, Nicola Colclough, and Brian Law

AstraZeneca Pharmaceuticals, Mereside, Alderley Park, Macclesfield SK10 4TG, U.K.

Received May 4, 2006

By identifying every pair of molecules that differ only by a particular, well-defined, structural transformation in a database of measured properties and computing the corresponding change in property, we obtain an overview of the effect that structural change has upon the property and set an expectation for what will happen when that transformation is applied elsewhere. The mean change indicates the expected magnitude of the change in the property and the number of cases in which the property increases give the probability that the structural transformation will cause the property to increase. Outliers indicate potential ways of avoiding the general trend. Comparing to changes in lipophilicity highlights structural transformations that have unusual effects, some of which can be explained by conformational changes. In this paper, we focus upon the effects on aqueous solubility, plasma protein binding and oral exposure of adding substituents to aromatic rings and methylating heteroatoms.

Introduction

The process of lead optimization (LO) involves the identification from within a lead series of the molecular entity with the most appropriate achievable balance of properties. The series of compounds that start this process, the "leads", will generally be the result of a search in which potency against the biological target of interest has been the key parameter. Other properties relevant to suitability as a drug candidate may have been examined but are unlikely to have been optimized. In particular, the structure−activity relationship (SAR) relating to the potency of the series may have been mapped, but the relationship with respect to other properties is unlikely to have been examined so carefully.

Within AstraZeneca, as with other pharmaceutical companies, databases of the properties relevant to any potential drug have been built up over many years. These should provide a rich mine of information that could guide the lead optimization process by identifying structural changes that are likely to be beneficial with respect to such properties. In this paper, we describe an approach that we have developed to mine these data and highlight how this can be deployed to guide molecular design and to aid decision making in the LO process.

Background

In the analysis presented here, the focus is upon aqueous solubility, binding to plasma proteins, and oral exposure in an in vivo model. Several methods to predict each of these properties have been presented before.[1−3] The majority involve the generation of a number of molecular descriptors and the identification of a mapping between these and the property of interest. This may be by way of a neural network, decision tree, partial least squares, multiple linear regression or other machine learning method, or a combination of such methods. Given a molecular structure, the property can be predicted (to a greater or lesser degree of certainty) and those compounds that are predicted to have a desired value of the property might be selected for synthesis. It is often not possible to know which structural changes might lead to the desired outcome, so all possible structures need to be envisaged to identify the best. Multiple linear regression and partial least-squares analysis can have the advantage of suggesting rules of thumb describing how a descriptor should change to achieve the desired change in the property. The structural changes that will achieve this are not identified. All of these methods are highly dependent upon the quality of the calculated descriptors and often involve extrapolation outside the chemical space upon which the models were built and where they are likely to be less reliable (or possibly to fail altogether). A variation on this is to use a model relating other measured data to a property of interest. This is often done to model in vivo effects using in vitro measurements but is exemplified by Yalkowsky and Valvani's work relating aqueous solubility to lipophilicity and melting point.[4] Other, more physical methods for prediction of solubility are available.[5]

A second way that data is frequently analyzed, particularly with regard to potency, is that a few matched pairs of molecules that differ only by a small structural change (a matched molecular pair) will be identified and the corresponding change in potency calculated.[6] This is most commonly used to breakdown SAR into manageable structural components and such matched structures can be designed systematically according to the Free−Wilson approach.[7] As demonstrated previously, the analysis can be used to study effects in ADME (absorption, distribution, metabolism, and excretion) and physical properties in a more generic fashion.[8] Sheridan et al. recently reported upon a data-mining tool that has been developed at Merck to identify all of the small structural transformations that have been tried in a particular series and their corresponding effects on activity.[9]

An example of a matched molecular pairs analysis in SAR is the pairs of molecules in Figure 1 taken from a publication from Pfizer concerning inhibitors of cyclin dependent kinase 4 (CDK4).[10] Although presented in that publication to show the effect of changing the amine substituent R′, the data provide a very clear profile of the effect of changing the second substituent (Br or Ac in **1** and **2** respectively) upon potency. The six sets

* Corresponding author. Phone: +44 1625 231853. Fax: +44 1625232693. E-mail: andrew.leach@astrazeneca.com.

10.1021/jm0605233 CCC: $33.50
Published on Web 10/20/2006

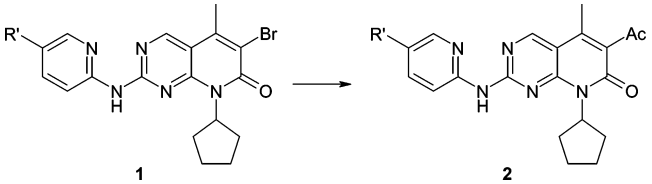


Figure 1. Pfizer CDK4 inhibitors: matched molecular pairs in which Br is exchanged for Ac have been made and tested.[10]

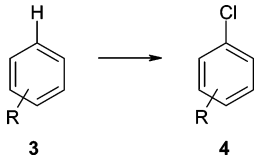


Figure 2. Matched molecular pairs corresponding to the addition of chlorine to an all-carbon aromatic ring.

Table 1. Matched Molecular Pairs (**1** and **2**) Demonstrating the Effect upon Potency (IC_{50}) of Changing Br to Ac in a Set of Compounds Illustrated in Figure 1[10]

	IC_{50} (μM)		pIC_{50}		
R′	Br	Ac	Br	Ac	pIC_{50}(Br) − pIC_{50}(Ac)
piperazine	0.16	0.011	6.80	7.96	+1.16
$(CH_3OCH_2CH_2)_2N$	1.1	0.051	5.96	7.29	+1.33
3,5-dimethylpiperazine	0.063	0.037	7.20	7.43	+0.23
N-methylpiperazine	0.136	0.005	6.87	8.30	+1.43
4-hydroxypiperidine	0.074	0.019	7.13	7.72	+0.59
morpholine	1.95	0.004	5.71	8.40	+2.69

Table 2. The Effect of Adding Chlorine to an Aromatic Ring on Aqueous Solubility for a Range of Simple Aromatic Compounds[11]

			log(solubility)[a]			
entry	substituent R	position relative to Cl in **4**	ArH (**3**)[11]	ArCl (**4**)[11]	change in[a] going from ArH to ArCl	*T* (°C)[11]
1	Me	4	−2.25	−3.02	−0.77	20
2	OH	2	−0.12	−0.86	−0.74	25
3	OH	3	0.01	−0.71	−0.72	20
4	OH	4	0.01	−0.78	−0.79	20
5	Ph	2	−4.27	−4.54	−0.27	25
6	Ph	3	−4.27	−4.94	−0.67	25
7	Ph	4	−4.27	−5.19	−0.92	25
8	NO2	2	−1.81	−2.55	−0.74	20
9	NO2	3	−1.81	−2.76	−0.95	20
10	NO2	4	−1.81	−2.74	−0.93	20
11	OMe	2	−1.92	−2.46	−0.54	25
12	OMe	3	−1.92	−2.78	−0.86	25
13	OMe	4	−1.92	−2.78	−0.86	25

[a] Solubility in water in M. This is the geometric mean of all of the solubility values given at the appropriate temperature in ref 11.

of data (other examples in which one of the data points was out of the range of the assay have been excluded), when regrouped as shown in Table 1, can be used to give a more balanced view of the effect of exchanging Br with Ac upon potency than any individual pair of structures might. Indeed, it is clear that, with a mean change in potency of 1.2 log units (pIC_{50}) and a standard error in that mean of 0.35, that exchanging Br with Ac in this position increases potency substantially and reliably, regardless of the nature of R′. However, the magnitude of the change does depend on R′, even though the two groups are rather distant from one another.

Similarly, the structures **3** and **4** (Figure 2) detailed in Table 2 corresponding to a set of matched molecular pairs involving addition of chlorine to simple aromatic compounds that all have aqueous solubility data available.[11] These compounds are listed, pairwise, in Table 2, and the logarithm of the aqueous solubility at either 20 or 25 °C is given for each compound. The table also gives the difference in these log(solubility) measurements for each set of molecular pairs. The matched molecular pairs involve a range of substituents with different electronic properties at different positions around the ring. If this set of substituents is sufficiently representative of all the substitution possibilities around the ring, then the mean change reflects the effect of chlorine addition independent of any interactions the chlorine may have with other substituents. The data in Table 2 reveal that the effect of adding chlorine is more or less independent of the solubility of ArH (**3**). For instance, contrasting entries 1 and 3, the solubility of ArH differs by over 2 orders of magnitude but the effect of adding chlorine is essentially identical in each case. Examining structures in a pairwise fashion reveals that addition of chlorine to an aromatic ring generally decreases the solubility, regardless of the position of attachment. The mean decrease in solubility is −0.75 log units, although these data are not all at one temperature and the solvent is unbuffered, so taking such a mean may not be appropriate in this instance.

Table 2 also demonstrates the value of the analysis for identifying interesting outliers. Entry 5, in which chlorine is added ortho to the phenyl substituent, is clearly unusual; it has a *Z*-score of 2.7 (the number of standard deviations away from the mean difference). The solubility decreases by a markedly smaller amount than in any of the other cases. Interestingly, the addition of a chlorine in this position changes the conformational profile about the Ar−Ph bond. According to the Cambridge Crystallographic Database,[12] in the absence of the chlorine (all four ortho positions bearing H), the angle between the planes of the two rings is generally between ±35°. When the chlorine is added, it is ∼55° (for details see Supporting Information). This increased twist is postulated to lead to less efficient crystal packing and hence a larger solubility than would be expected because the solid state is less stable.

Matched molecular pairs such as those above are usually identified by hand, making the process very time-consuming and prone to a lack of objectivity. Notably, pairs can be chosen selectively to support a case while other pairs go unreported. The number of pairs that are identified will be low and hence likely to lack statistical significance. The method has the advantage of involving only measured data, with no extrapolation to unmeasured compounds. If the set of pairs chosen were to be large and diverse enough, the computed effect of the structural change should set the expectation for what will happen in any series.

A suite of programs that can be used to identify all the pairs of molecules that fit a particular structural change within a set of molecules has therefore been developed at AstraZeneca. The set of molecules in each case consists of all those compounds with entries in a database of measured molecular properties. Effects on ADME and physical properties have been our focus due to the value in understanding their link to structure in the lead optimization process. In this paper, we discuss three properties: aqueous solubility, plasma protein binding, and oral exposure. Such properties are relevant to a drug-hunting program, regardless of the target, so the analysis should be very widely applicable. The statistical analysis presented here yields the probability that a particular structural change will push a property in the desired direction and predicts the likely magnitude of such a change and the variability in that magnitude. The most general view of the effect of that structural change on the particular molecular property available from the dataset can therefore be assessed.

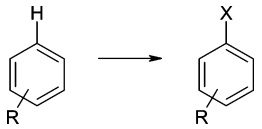


Figure 3. Matched molecular pairs for substitution on an aromatic ring detailed in Tables 3−5.

Aqueous solubility is a key parameter for drugs, as it fundamentally limits the amount of compound that can be present in solution in the gut and in the circulation.[11,13] Solubility can limit the bioavailability of a compound and lead to high patient-to-patient and fed-to-fasted variation. The FDA regulations concerning oral medications require more investigation of low solubility compounds.[14] Plasma protein binding limits the amount of compound that is free in solution in the plasma and influences the volume of distribution and clearance of a compound.[15] Binding can be to any of the many proteins that are present in the plasma. The most significant of these groups of proteins is the albumins,[16] but others such as α-1-acid glycoproteins[17] are also present and can contribute to the reduced level of free compound in solution. The presence of a drug in the plasma at detectable levels after oral dosing is a key requirement; this can be quantitated from the area underneath the concentration versus time plot when plasma concentration is monitored. This oral exposure is a composite property depending upon many other parameters, such as solubility, permeability, efflux potential, metabolism, excretion, and plasma protein binding.

Results and Discussion

The analysis will be illustrated by two sets of structural changes: the addition of substituents to an aromatic ring and the methylation of heteroatoms. Their effect on three pharmaceutically relevant properties—aqueous solubility, rat plasma protein binding, and oral exposure in rats—has been studied.

(a) Substituents on Six-Membered All-Carbon Aromatic Rings. The AstraZeneca (Alderley Park) databases of aqueous solubility measurements, rat plasma protein binding measurements, and rat oral exposure measurements were investigated to find all occurrences of pairs fitting the description in Figure 3 for a small range of X. In this case, compounds were not limited to single aromatic rings, such that the substituents on the six-membered ring in fused systems, for example indoles, would also be included. This is one of the subtleties of the analysis; it is essential that the results are presented in such a way that the reader is clear how narrowly defined the scope of the pairs is. In our programs for finding matched molecular pairs, the pair members are defined by SMARTS and hence their chemical structure can be tightly controlled. After finding all pairs, the corresponding solubilities, K_1 values for rat plasma protein binding, and area under the curve values for rat oral exposure were added, and the difference in the log of the solubilities, difference in the log of K_1, and the difference in the log of the area under the curve [log(AUC)] were computed for each pair. Each of these properties is normally distributed on a logarithmic scale and for each set of pairs, the distribution of the differences is approximately normal. The mean changes in each property for each set of pairs is computed, along with the corresponding standard deviation and standard error of the mean. Furthermore, the percentage of cases in which addition of the substituent causes an increase in each property (regardless of by how much) has been computed. The 95% confidence interval on this percentage is computed from the binomial probability function.[18] The total number of pairs in each case is also listed. Experience with this kind of analysis has suggested that there should be at least 20 pairs for the set to be useful. The analysis relies upon the set of pairs being representative. However, sometimes a particular transformation will be represented exclusively, or almost exclusively, by pairs from one particular series and the set may not be representative of the general trend. Only well-populated, diverse compound sets are given in this paper. Table 3 gives the data for solubility, Table 4 for plasma protein binding, and Table 5 for oral exposure.

Table 3. The Effect of Adding Substituents to Aromatic Rings upon Aqueous Solubility

	change in log(solubility)					
X	mean[a]	SD[b]	SEM[c]	no. of pairs	% in which solubility increases[d]	mean change in clogP[e]
F	−0.22	0.69	0.03	711	34 (30−37)	+0.13
Cl	−0.67	0.69	0.04	326	14 (11−18)	+0.69
Br	−1.10	0.67	0.10	44	2 (0−12)	+0.89
CF_3	−0.81	0.71	0.09	69	17 (9−28)	+0.95
Me	−0.21	0.71	0.06	166	33 (26−41)	+0.43
OMe	−0.11	0.68	0.04	297	42 (37−48)	−0.01
CN	−0.26	0.87	0.09	104	36 (26−46)	−0.38
OH	+0.07	1.00	0.18	32	56 (38−74)	−0.57
SO_2Me	+0.26	0.65	0.12	28	71 (51−87)	−1.51

[a] The mean value of log(solubility) for ArX − log(solubility) for ArH, where solubility is in M. [b] The standard deviation of the distribution of the values of log(solubility) for ArX − log(solubility) for ArH. [c] The standard error of the mean in column 2. [d] The percent of cases in which the solubility increases, no matter by how much; the values in parentheses are the 95% confidence limits in this value (see ref 18). [e] The mean of the distribution of values of clogP for ArX − clogP for ArH (see ref 19).

Table 4. Effect of Adding Substituents to Aromatic Rings upon Rat Plasma Protein Binding

	change in log(K_1)					
X	mean[a]	SD[b]	SEM[c]	no. of pairs	% in which K_1 increases[d]	mean change in clogP[e]
F	+0.15	0.29	0.01	467	77 (73−81)	+0.13
Cl	+0.54	0.35	0.02	242	95 (92−98)	+0.65
Br	+0.78	0.45	0.07	39	97 (87−100)	+0.88
CF_3	+0.60	0.45	0.07	46	89 (76−96)	+1.00
Me	+0.22	0.35	0.03	115	79 (71−86)	+0.42
OMe	+0.02	0.41	0.03	171	58 (50−65)	−0.06
CN	−0.06	0.40	0.04	60	38 (26−52)	−0.38
OH	−0.24	0.36	0.06	31	29 (14−48)	−0.53
SO_2Me	−0.55	0.37	0.08	20	0 (0−14)	−1.43

[a] The mean value of log(K_1) for ArX − log(K_1) for ArH, where K_1 is the association constant for rat plasma proteins, concentrations being measured in M. [b] The standard deviation of the distribution of the values of log(K_1) for ArX − log(K_1) for ArH. [c] The standard error of the mean in column 2. [d] The percent of cases in which the plasma protein binding increases, no matter by how much; the values in parentheses are the 95% confidence limits in this value (see ref 18). [e] The mean of the distribution of values of clogP for ArX − clogP for ArH (see ref 19).

Table 3 reveals that many effects upon aqueous solubility that might be expected are indeed observed. For instance, the addition of the heavier halogens is detrimental to solubility. More than that, it provides a numerical estimate for how large their effect upon solubility will be. It reveals that adding bromine to an aromatic ring leads, on average, to over an order of magnitude reduction of aqueous solubility. In terms of making decisions for synthesis, it reveals that adding bromine led to a decrease in solubility in 43 of the 44 cases (98%) where the data are available in our database. This can be approximately equated to the probability of seeing a decrease in solubility when bromine is added to the aromatic ring of a different compound. Thus, if solubility is a major issue in a drug-hunting project, then addition of bromine should be a very low synthetic priority, and if bromine is already present, its removal should be considered.

Table 5. Effect of Adding Substituents to Aromatic Rings upon Oral Exposure in an in Vivo Rat Model

X	change in log(AUC) mean[a]	SD[b]	SEM[c]	no. of pairs	% in which log(AUC) increases[d]	mean change in clogP[e]
F	+0.09	0.65	0.03	551	55 (51−60)	+0.14
Cl	−0.03	0.74	0.04	359	44 (39−49)	+0.63
Br	+0.02	0.75	0.11	49	53 (38−67)	+0.89
CF_3	−0.03	0.97	0.12	71	48 (36−60)	+0.97
Me	−0.18	0.74	0.06	176	39 (32−47)	+0.47
OMe	−0.18	0.84	0.06	226	42 (36−49)	−0.07
CN	−0.04	0.99	0.11	83	41 (30−52)	−0.39
OH	−0.19	0.90	0.12	54	48 (34−62)	−0.54
SO_2Me	−0.11	0.95	0.13	51	53 (38−67)	−1.36

[a] The mean value of log(AUC) for ArX − log(AUC) for ArH, where AUC is the dose-normalized area under the curve in a rat oral exposure assay in (h μg/mL)/(mg/kg). [b] The standard deviation of the distribution of the values of log(AUC) for ArX − log(AUC) for ArH. [c] The standard error of the mean in column 2. [d] The percent of cases in which the AUC increases, no matter by how much; the values in parentheses are the 95% confidence limits in this value (see ref 18). [e] The mean of the distribution of values of clogP for ArX − clogP for ArH (see ref 19).

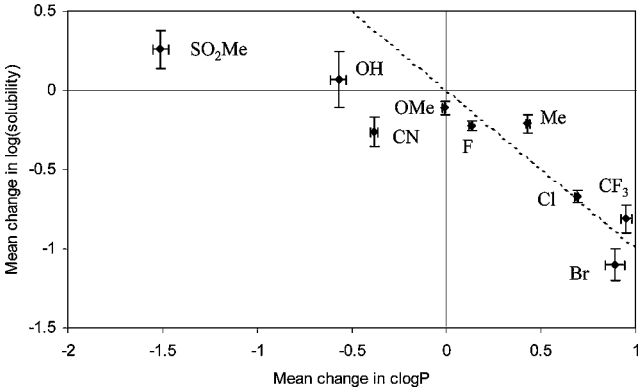


Figure 4. The mean change in log(solubility) plotted against the mean change in clogP for a set of matched molecular pairs involving the addition of substituents to a six-membered all-carbon aromatic ring. Error bars correspond to a single standard error in each mean. A line of best fit is not plotted because of the stratification into groups involving changes in the hydrogen bond acceptor or donor count and those that do not. The dashed line is a line of slope −1 through the origin.

To identify whether the effect on solubility determined for each transformation is merely a reflection of the change in lipophilicity, the corresponding change in clogP for all of the matched molecular pairs was computed.[19] In Figure 4, the mean change in solubility is plotted against the mean change in clogP. The error bars are the standard error of each mean. According to the analysis of Yalkowsky and Valvani, the solubility should be given by[4,21]

$$\log S_W \approx -\log P - 0.01 \times \text{mp} + 1.05 \qquad (1)$$

and hence changes in solubility are given by[4]

$$\Delta\log S_W \approx -\Delta\log P - 0.01 \times \Delta\text{mp} \qquad (2)$$

This is in effect a lipophilicity dependence with slope −1 and a term dependent on the solid state. Figure 4 is consistent with this model. There is a trend for increases in lipophilicity to correspond to decreases in solubility; notably, the changes for OMe, F, Me, Cl and CF_3 do fall around a line of slope ca. −1, such as the one plotted in Figure 4, which goes through the origin. Furthermore, introduction of groups that involve hydrogen-bond acceptors causes a larger decrease in solubility or smaller increase than their effect on lipophilicity would suggest (the same is also true of donors as detailed in section b). These groups have the potential to be involved in interactions in the solid state that are stronger than any solvation they may engender in the solution state. This is consistent with eq 2, because these structural transformations are expected to lead to changes in the melting point of these compounds. An interesting exception to this is the addition of OMe, reflecting the exceptionally poor hydrogen-bond-acceptor qualities of anisole type ethers.[20] This contrasts with the better acceptor qualities of dialkyl ethers, which lead to their marked effects on solubility observed in other matched pairs analyses not reported here.

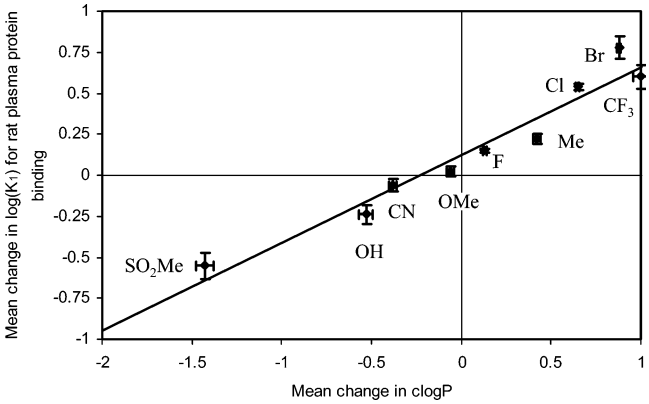


Figure 5. The mean change in log(K_1) for rat plasma protein binding plotted against the mean change in clogP for a set of matched molecular pairs involving the addition of substituents to a six-membered all carbon aromatic ring. Error bars correspond to a single standard error in each mean. The dark line is the line of best fit ($R^2 = 0.94$).

In contrast to the picture for aqueous solubility, the plot of changes in rat plasma protein binding against changes in clogP in Figure 5 (from data in Table 4) reveals that lipophilicity accounts for the general trend seen in changes in the association constant for the addition of various substituents to aromatic rings. Although it is common within AstraZeneca to measure plasma protein binding in rat plasma, particularly for drug discovery applications, the property that may be of more interest is the plasma protein binding in man, which is used in predictions of the human pharmacokinetics.[22] Although the AstraZeneca database is very much smaller for human plasma protein binding, the data for the same sets of matched pairs that are sufficiently represented have been identified. In Figure 6, the mean change in human plasma protein binding is plotted against the mean change in rat plasma protein binding (human data is given in the Supporting Information). The error bars correspond to a single standard error of the mean. The plot reveals that changes in rat plasma protein binding in general translate very well to the corresponding change in human plasma protein binding. This suggests that rat plasma protein binding is an appropriate surrogate for the human equivalent to drive medicinal chemistry against, as well as suggesting that the matched molecular pairs analysis described here is relevant to human as well as rat plasma protein binding.

The final property investigated for this set of structural transformations is oral exposure as measured by the area under the curve (normalized for dose) in a high-throughput blood level assay in rats. The mean difference in log(AUC) has been computed for each matched molecular pair and is listed in Table 5. In Figure 7, the change is plotted against the corresponding change in clogP. The error bars are the standard errors of each mean. This plot shows that for the addition of Me or OMe to an aromatic ring there is a statistically significant decrease in blood levels and for the addition of F there is a small but statistically

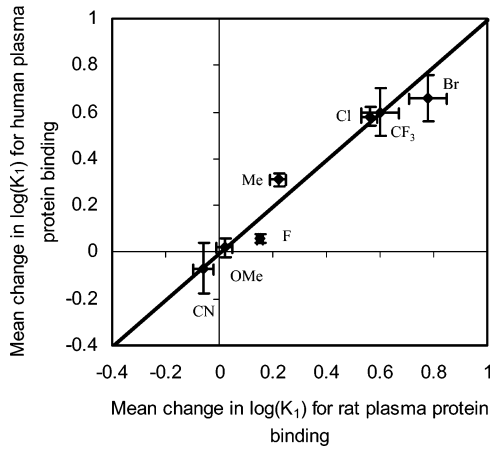


Figure 6. The mean change in log(K_1) for human plasma protein binding against the corresponding change for rat. The error bars correspond to a single standard error in each mean and the dark line is the line of equality.

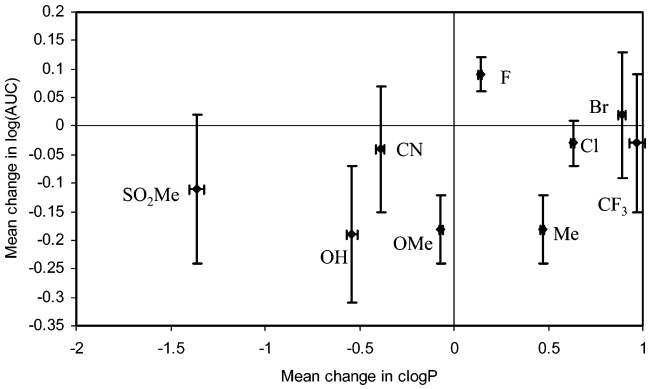


Figure 7. The mean change in log(AUC) for rat oral exposure plotted against the mean change in clogP for a set of matched molecular pairs involving the addition of substituents to a six-membered all-carbon aromatic ring. Error bars correspond to a single standard error in each mean.

significant increase in blood level. However, in all cases there is a very broad variation in the effect, as indicated by the large standard deviations and the fact that most of the means are not statistically distinct from a change of 0 at 95% confidence. The probabilities are significant in four cases that indicate that adding F is likely to be beneficial in increasing the AUC whereas adding Cl, Me, or OMe will be detrimental. This property is more suitable for a study localized to a series of interest, because the position at which the substituent is added clearly influences the effect it has. This is frequently the case for potency too, because substituents will be involved in interactions with the target protein that will be very dependent upon their position in three dimensions. In the present case of oral exposure, the ability of substituents to block metabolism, for instance, will be highly dependent upon the detailed structure of the molecule.

Taken together, Tables 3−5 indicate, for instance, that adding a fluorine to an aromatic ring will decrease solubility (66% chance), increase plasma protein binding (77% chance), and increase oral exposure (55% chance). In this case, all of these probabilities are on sufficiently large datasets that they are statistically distinct from 50% at the 95% confidence interval. Addition of fluorine might be worth considering in a series that is not too highly protein bound and in which oral exposure is lacking. Similarly, if solubility were an issue, it might be worth removing any chlorine or bromine substituents or adding a methyl sulfone substituent.

Table 6. Effect of Interchanging Halogen Substituents on Six-Membered Aromatic Rings from Ar−X_1 to Ar−X_2 upon Aqueous Solubility

		change in log(solubility)		no. of
X_1	X_2	mean[a]	SEM[b]	pairs
F	Cl	−0.49	0.04	231
F	Br	−0.74	0.12	28
Cl	Br	−0.21	0.07	47

[a] The mean value of log(solubility) for Ar−X_2 − log(solubility) for Ar−X_1, where solubility is in M. [b] The standard error of the mean in column 3.

Table 7. Effect of Interchanging Halogen Substituents on Six-Membered Aromatic Rings from Ar−X_1 to Ar−X_2 upon Plasma Protein Binding

		change in log(K_1)		no. of
X_1	X_2	mean[a]	SEM[b]	pairs
F	Cl	+0.43	0.02	195
F	Br	+0.60	0.06	41
Cl	Br	+0.15	0.02	59

[a] The mean value of log(K_1) for Ar−X_2 − log(K_1) for Ar−X_1, where K_1 is the association constant for rat plasma proteins, the concentrations being measured in M. [b] The standard error of the mean in column 3.

Table 8. Effect of Interchanging Halogen Substituents on Six-Membered Aromatic Rings from Ar−X_1 to Ar−X_2 upon Oral Exposure in a Rat in Vivo Model

		change in log(AUC)		no. of
X_1	X_2	mean[a]	SEM[b]	pairs
F	Cl	−0.01	0.03	311
F	Br	−0.13	0.09	45
Cl	Br	−0.11	0.05	72

[a] The mean value of log(AUC) for ArX_2 − log(AUC) for ArX_1, where AUC is the dose-normalized area under the curve in a rat oral exposure assay in (h μg/mL)/(mg/kg). [b] The standard error of the mean in column 3.

Often, there will be a substituent present in a compound on an aromatic ring and it will be of interest what the effect of changing that substituent to something else will be rather than of simply removing it. To test that the mean changes in Tables 3−5 are relevant to this, the effect of interchanging the three halogens has been probed for all three properties described so far. This is detailed in Tables 6−8 for each property in turn. Although there is some overlap between the compounds involved in deriving the initial set of matched pairs and those in the set for halogen interchange, the two sets are derived independently of one another. If they are consistent, it gives an indication that the effects being studied are indeed independent of the structure of the remainder of the molecule. The data for solubility, plasma protein binding and oral exposure in Tables 3−8 reveal that all of the step change means are within two standard errors of one another whether derived for substituent addition or substituent interchange. Hence, Tables 3−5 can be used to predict the effect for substituent interchange from the difference in the means for substituent addition. Furthermore, the means thus derived can be related to a probability using the plot of the percent of cases leading to an increase against the mean change in Figure 8 for aqueous solubility, Figure 9 for plasma protein binding, and Figure 10 for oral exposure. This is illustrated with arrows in the plot for aqueous solubility in Figure 8; a mean change in log(solubility) of −0.7 corresponds to a 15−20% chance that the corresponding transformation will increase solubility. This set of transformations does not provide a good dataset for the oral exposure assay for this analysis and leads to the flat line at ∼50% in Figure 10.

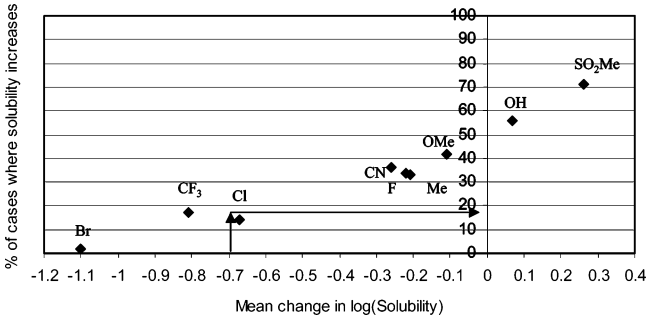


Figure 8. The percent of cases in which an increase in log(solubility) is observed compared to the mean change in log(solubility).

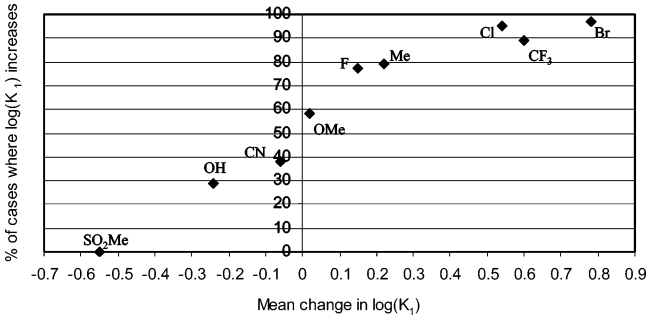


Figure 9. The percent of cases in which an increase in log(K_1) for plasma protein binding is observed compared to the mean change in log(K_1).

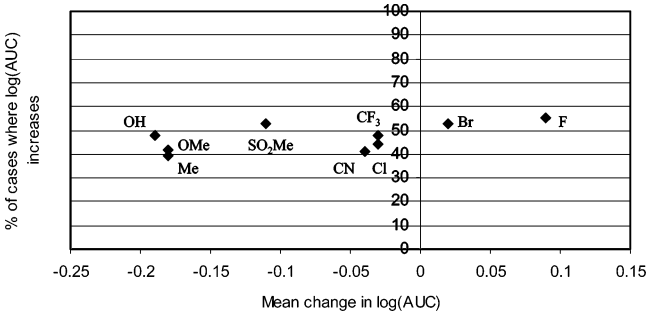


Figure 10. The percent of cases in which an increase in log(AUC) for oral exposure is observed compared to the mean change in log(AUC).

The discussion above indicates that, on average, the addition of different substituents to an aromatic ring will cause the properties to change to different degrees. This need not mean that for any one compound that the effect of adding the different substituents can be distinguished. To probe this, the sets of molecular triplets in which F and Cl have been added to the same position on an aromatic ring have been found and are presented side by side in the Supporting Information. Other substituent interchanges were less well represented in the datasets. The solubility of the chloride (ArCl) and the fluoride (ArF) is plotted against the corresponding solubility of the unsubstituted compound (ArH) in Figure 11. The corresponding plots for plasma protein binding and oral exposure are in Figures 12 and 13, respectively. As might have been expected from the data in Table 3, adding chlorine to an aromatic ring does, in general, decrease the solubility by more than adding fluorine in the same place. Indeed, in 78% of cases where this has been done, the chloride is less soluble. Similarly, the trend for plasma protein binding is clear in Figure 12, and adding chlorine leads to a more highly protein bound compound than fluorine in 90% of the cases where the corresponding compounds have been made. The indistinct trends in the oral exposure are reflected

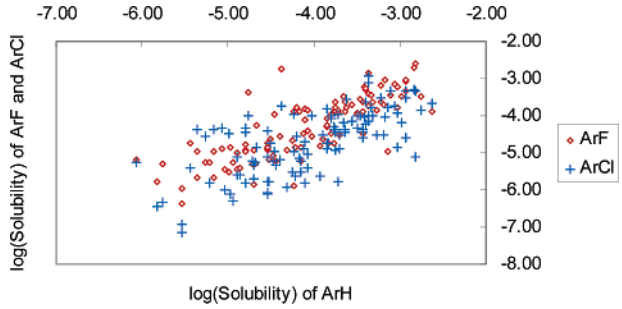


Figure 11. The log(solubility) of corresponding aromatic chlorides and fluorides compared to the log(solubility) of the parent aromatic compound.

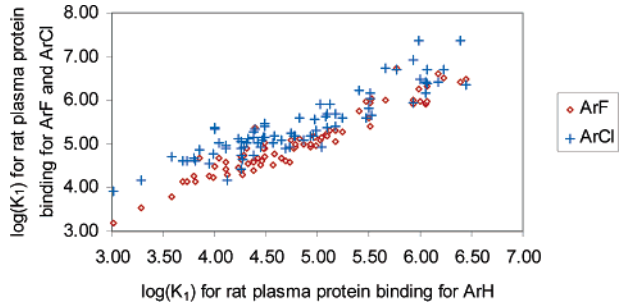


Figure 12. The log(K_1) for plasma protein binding of corresponding aromatic chlorides and fluorides compared to the log(K_1) of the parent aromatic compound.

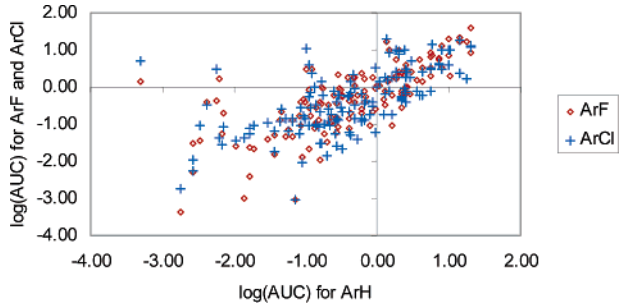


Figure 13. The log(AUC) for rat oral exposure of corresponding aromatic chlorides and fluorides compared to the log(AUC) of the parent aromatic compound.

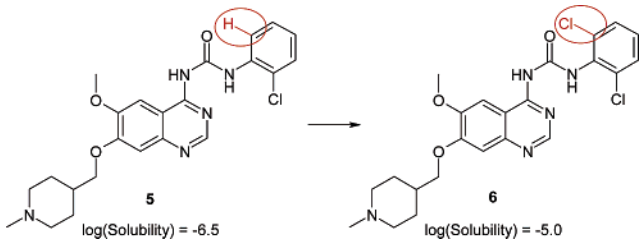


Figure 14. An unusual change in solubility: adding chlorine to **5** causes an increase in solubility.

in the overlaid distributions in Figure 13, and the chloride has a lower AUC in 57% of cases.

The analyses described so far serve a further function to that already suggested. By quantitating what may be expected from the addition of a particular substituent to an aromatic ring, it is possible to identify those transformations that lead to an unexpected result. For instance, the transformation from **5** to **6** in which adding chlorine actually causes an increase in solubility by 1.5 log units is far from expectation (*Z*-score of 3.1) (Figure 14). This can be understood in terms of a conformational effect constraining the urea group to be orthogonal to the aromatic ring in **6**, whereas more planar conformations that may stack

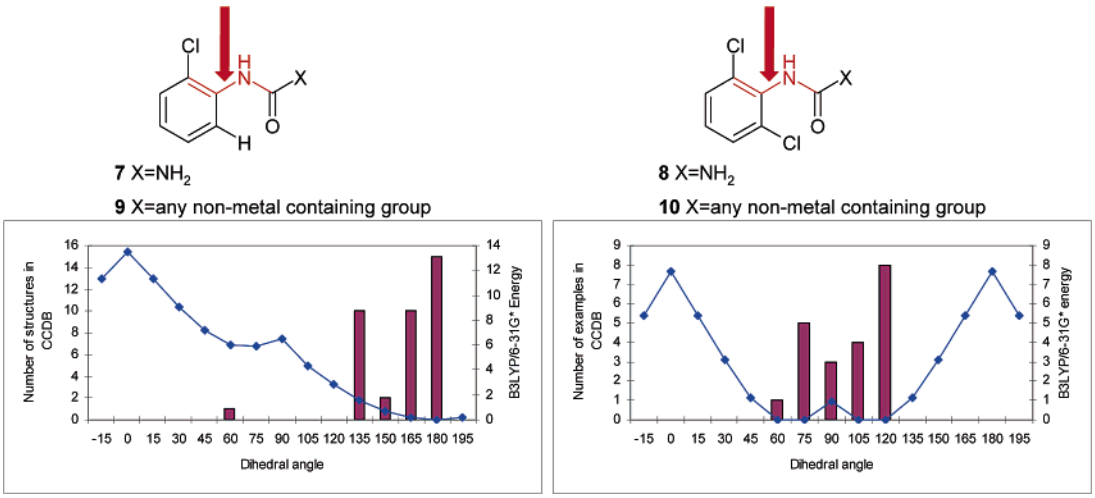


Figure 15. The conformational preferences of compounds with the substructures indicated. The *x*-axis on each plot gives values of the torsion angle indicated in red. The curve is the energy computed by B3LYP/6-31G* for **7** (left-hand side) and **8** (right-hand side) for structures in which the dihedral angle is constrained to the angle shown. The overlaying bars correspond to the number of structures containing **9** (left-hand side) or **10** (right-hand side) with a torsion angle falling within a bin centered at each value on the *x*-axis.

better in the solid state are permitted in **5**. Figure 15 shows how the energy of the model compounds **7** and **8** varies with the torsion highlighted twisting the urea out of plane as computed at the B3LYP/6-31G*[a] level (constrained to the angle indicated at each point and with the conformation shown about the urea C−N bond).[2,23,24] Overlaid with this is the number of compounds with the substructure shown as **9** and **10** that have torsions within the bin centered on each torsion angle that are observed in the Cambridge crystallographic database.[12] It can be seen that these two measures agree well and that low-energy structures are well-represented in the Cambridge database. The structures with just one *o*-chlorine are expected to be planar and those with two *o*-chlorines are expected to be almost orthogonal. The full data set is reported in the Supporting Information.

The dramatic effect upon solubility of adding chlorine to **5** to get **6** highlights the importance of conformational effects in transcending expectation. To investigate the generality of this effect, the small number of pairs in which a substituent is added ortho to an anilide already bearing a single heavy atom substituent (F, Cl, Br, Me, OH, etc.) in the other ortho position, illustrated by **11** and **12** in Figure 16, have been located and characterized by their *Z*-score for all three properties. There are three examples for X = F in the solubility dataset with a mean *Z*-score for the change in log(solubility) of +1.49 (range = +0.8 to +2.6), one example for X = Cl with a *Z*-score of +3.10, and three examples for X = Me with a mean *Z*-score of +1.56 (range = +0.59 to +2.41). These examples are clearly outliers in the positive direction, which is to say that the solubility does not decrease by as much (or increases) as is usual for the addition of each of these substituents. In the protein binding matched pairs, there is one example for X = Cl with a *Z*-score of −6.01 and one example for X = Me with a *Z*-score of −5.63. These massive outliers indicate that protein binding does not increase by anywhere near as much as expected for the general case of adding these substituents. Finally, in the oral exposure data, there is one example for X = F with a *Z*-score of 0.08 and one for X = Me with a *Z*-score of −0.10. The effect on AUC is clearly unexceptional.

[a] Abbreviations: B3LYP/6-31G*, density functional theory method employing Becke's 3 parameter exchange functional and the Lee, Yang, Parr correlation functional and the 6-31G* basis set; CCDB, Cambridge Crystallographic Database; SMILES, Simplified Molecular Input Line Entry System; SMARTS, SMILES Arbitrary Target Specification.

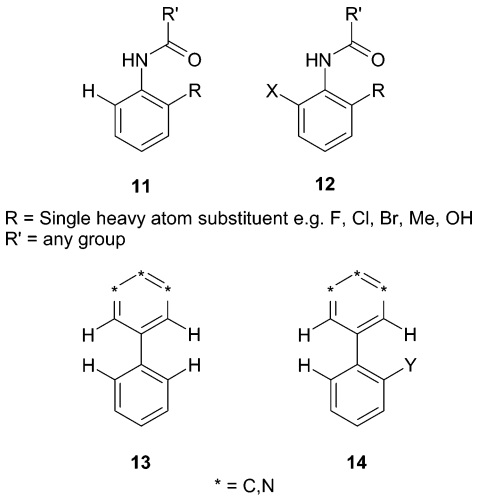


Figure 16. Structural changes that lead to significant changes in conformation.

Entry 5 in Table 2 highlighted the effect of adding substituents ortho to a biphenyl-type linkage, and this was rationalized as a conformational effect. The generality of this effect has been studied by finding all matched pairs resembling **13** and **14** (Figure 16). In the aqueous solubility matched pairs set, there are 13 examples for Y = F with a mean *Z*-score of +0.54 (range = −0.44 to +1.45) and there are five examples for Y = Cl with a mean *Z*-score of +1.93 (range = +1.34 to +2.42). In the plasma protein binding matched pairs, there are nine examples for Y = F with mean *Z*-score of −0.80 (range = −3.69 to +0.01) and two examples for Y = Cl with a mean *Z*-score of −0.60 (range = −0.81 to −0.39). In the oral exposure dataset, there are 12 examples for Y = F with a mean *Z*-score of −0.34 (range = −2.25 to +1.51) and one example for Y = Cl with a *Z*-score of 1.01. The pairs corresponding to **11** and **12** as well as **13** and **14** indicate that when a conformational effect causes compounds to prefer to be less planar, solubility tends to increase more than expected (or decrease less than expected), the extent of plasma protein binding tends to increase less than expected, and AUC does not vary in a predictable fashion. A visual inspection of all of the outliers (|*Z*-score| > 2) in each distribution indicates that in the case of solubility, the majority of the outliers could have a contribution from a conformational effect, whereas the distinction was not so clear

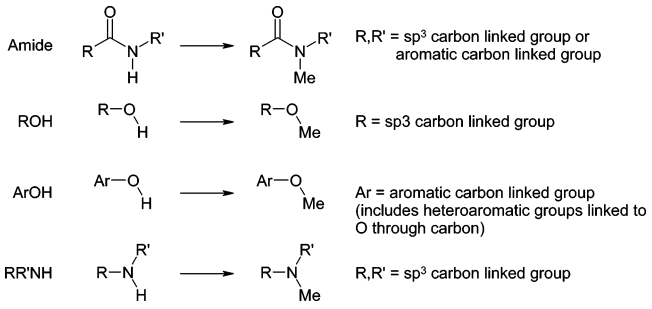


Figure 17. A number of heteroatom methylations for which matched molecular pairs have been identified.

Table 9. The Effect of Methylating Some Heteroatoms upon Aqueous Solubility

	change in log(solubility)					
X−H	mean[a]	SD[b]	SEM[c]	no. of pairs	% in which solubility increases[d]	mean change in clogP[e]
amide	+0.64	0.73	0.06	142	79 (71−85)	+0.31
ROH	−0.01	0.68	0.05	203	44 (37−51)	+0.46
ArOH	−0.22	0.84	0.20	17	41 (18−67)	+0.10
RR′NH	−0.32	0.69	0.07	107	34 (25−43)	+0.30

[a] The mean value of log(solubility) for X−Me − log(solubility) for X−H, where solubility is in M. [b] The standard deviation of the distribution of the values of log(solubility) for X−Me − log(solubility) for X−H. [c] The standard error of the mean in column 2. [d] The percent of cases in which the solubility increases, no matter by how much; the values in parentheses are the 95% confidence limits in this value (see ref 18). [e] The mean of the distribution of values of clogP for X−Me − clogP for X−H (see ref 19).

in the protein binding pairs, and as might be expected, there is little distinction in the oral exposure matched pairs. Most of the outliers are assumed to arise due to the experimental variation that should be expected in such large datasets.

(b) Methylation of Heteroatoms. By way of a second illustration of the type of structural transformation whose effect upon properties may be considered, the effect of methylating heteroatoms in a small number of functional groups has been investigated. These transformations are described in Figure 17 and the effects on the same three properties discussed above are given in Tables 9−11.

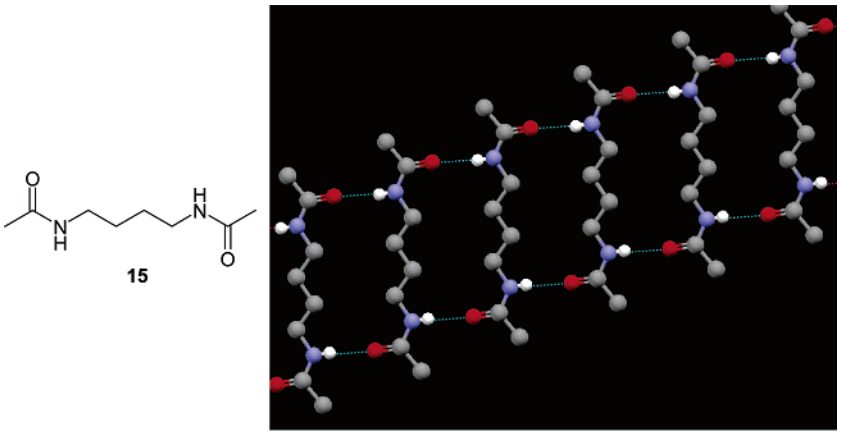
A number of the effects upon solubility in Table 9 may be surprising. The first is that methylation of alkyl alcohols has almost no effect upon solubility on average. In this transformation, the lipophilicity is being increased, which should diminish solubility (principally an aqueous phase effect), whereas removal of a hydrogen-bond donor should increase solubility (principally a solid state effect). These two effects cancel out overall. The most surprising effect is the substantial increase in solubility predicted for the methylation of amides. Here, the removal of the donor must completely outweigh any effect of the increase in lipophilicity. Presumably, this reflects the ability of amides with a free N−H to provide ideal donor−acceptor complementarity in the solid state, which can lead to infinite ladders of hydrogen bonds, as exemplified by the crystal structure of **15** in Figure 18. Furthermore, N-methylation is expected to remove the preference for the conformation in which R and R′ (in Figure 17) are antiperiplanar to one another that is strongly favored by secondary amides. The *N*-methylamides populate both conformations about the amide bond in solution and so will suffer a larger entropic loss upon rigidification in the crystal lattice.

Table 10. Effect of Methylating Some Heteroatoms upon Rat Plasma Protein Binding.

	change in log(K_1)					
X−H	mean[a]	SD[b]	SEM[c]	no. of pairs	% in which K_1 increases[d]	mean change in clogP[e]
amide	−0.23	0.31	0.03	88	22 (14−32)	+0.34
ROH	+0.14	0.26	0.02	154	74 (66−81)	+0.42
ArOH	+0.09	0.44	0.11	16	63 (35−85)	+0.21
RR′NH	+0.17	0.33	0.03	104	69 (59−78)	+0.21

[a] The mean value of log(K_1) for X−Me − log(K_1) for X−H, where K_1 is the association constant for rat plasma proteins concentrations being measured in M. [b] The standard deviation of the distribution of the values of log(K_1) for X−Me − log(K_1) for X−H. [c] The standard error of the mean in column 2. [d] The percent of cases in which the protein binding increases, no matter by how much; the values in parentheses are the 95% confidence limits in this value (see ref 18). [e] The mean of the distribution of values of clogP for X−Me − clogP for X−H (see ref 19).

The effect of methylations upon protein binding is given in Table 10 and illustrated with the plot against mean change in clogP in the Supporting Information. Once again, amide methylation is an obvious outlier. Amide methylation decreases the extent of plasma protein binding despite the apparent added lipophilicity. This may reflect an error in clogP or the ability of amides with free N−Hs to interact well with other such amides (as discussed for their solubility), like those making up the plasma proteins. Finally, in contrast to the lack of consistent effects upon oral exposure observed for the addition of substituents to aromatic rings, there is a marked effect upon oral exposure when particular groups are methylated, as shown in Table 11 and in the plot against clogP in the Supporting Information. Notably, despite enhancing solubility and likely related to the decrease in plasma protein binding (which should increase overall metabolic clearance), amide methylation decreases log(AUC) on average. Methylation of phenolic hy-



15

Figure 18. The ladder of hydrogen bonds for *N*-(4-acetylaminobutyl)acetamide, **15** (Cambridge database code ABAWOQ).[12,25]

Table 11. Effect of Methylating Some Heteroatoms upon Rat Oral Exposure

	change in log(AUC)					
X−H	mean[a]	SD[b]	SEM[c]	no. of pairs	% in which log(AUC) increases[d]	mean change in clogP[e]
amide	−0.28	0.77	0.07	113	33 (24−42)	+0.20
ROH	+0.06	0.84	0.07	161	43 (35−51)	+0.53
ArOH	+0.59	0.87	0.21	18	78 (52−94)	+0.49
RR′NH	+0.16	1.01	0.15	46	63 (48−77)	+0.45

[a] The mean value of log(AUC) for X−Me − log(AUC) for X−H, where AUC is the dose-normalized area under the curve in a rat oral exposure assay in (h·μg/mL)/(mg/kg). [b] The standard deviation of the distribution of the values of log(AUC) for X−Me − log(AUC) for X−H. [c] The standard error of the mean in column 2. [d] The percent of cases in which the AUC increases, no matter by how much; the values in parentheses are the 95% confidence limits in this value (see ref 18). [e] The mean of the distribution of values of clogP for X−Me − clogP for X−H (see ref 19).

droxyls increases log(AUC), accompanying a small increase in plasma protein binding.

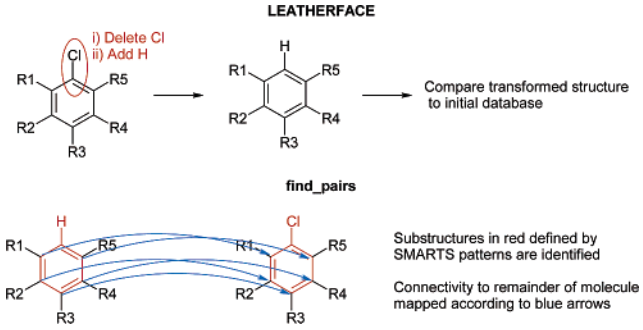
Conclusions

The above results indicate that building up a database of structural transformations and their likely effect upon a broad range of properties can provide a useful tool for the optimization of pharmaceutically relevant properties. Notably, during the lead optimization phase of a drug discovery program, relatively conservative structural perturbations, such as those described above, may be considered around a particular molecular core. The examples illustrate that a global matched molecular pairs analysis can be used to guide molecular design as it naturally suggests structural transformations that may have the desired effect upon a particular property or across a range of properties. The method can be applied to other structural transformations, many of which will be of more interest to medicinal chemists. The method can be used as a tool to test many of the "rules of thumb" that abound within medicinal chemistry. By analyzing the data in a categorical sense (increase or decrease), an approximate probability that the structural change will cause the property to move in one direction or another can be computed, assuming that the effects observed previously can be projected forward to different chemotypes. This analysis does rely upon a diverse data set being available.

Although here the focus has been upon three properties (four if clogP is included), any property can be treated just as well, although the larger the experimental variation, the more difficult it is to identify statistically significant effects. The data presented here show that analysis of properties such as solubility that depend on a limited number of equilibria or rates are more likely to present clear trends. In more complex properties, such as oral exposure, the structural dependency of the many competing processes may confound the identification of these trends. The derived information should be applicable regardless of the particular biological target of the drug-hunting project. If the mean change is to be computed, the property should be on a scale on which it is normally distributed. If very clear trends are identified, the analysis can have the considerable benefit of reducing the need for some measurements, which is particularly important where in vivo experiments can be avoided.

In order for the analysis to be useful, it is not essential that the cause of the effects be understood. It is often enough just to know what effect a structural change is likely to have. However, as the brief aside concerning the solubility effect of the methylation of amides indicates, the analysis of these effects can provide insight.

Scheme 1. Overview of the Process for Identifying Matched Molecular Pairs with Either the Molecular Editor LEATHERFACE (top) or the find_pairs Program (bottom)



By quantitating the effect that particular transformations are expected to have, the analysis permits unusual effects to be identified. These fall into two categories: the first is a structural change that has a surprising effect, such as the effect of methylation of an amide upon solubility, and the second is an unusual effect within a set of matched molecular pairs, such as the effect of going from **5** to **6** upon solubility. Both of these can provide valuable ideas for compound optimization.

Methods

Identifying an appropriate set of matched molecular pairs requires informatics tools that are capable of finding pairs of molecules within a set that are related to one another by a defined structural difference. Our database in this case consists of a set of molecules represented by their SMILES and a measured property for each. We have two different approaches for identifying matched pairs in these databases. The first involves the in-house program LEATHERFACE.[8] This is a molecular editor that can be used to transform the SMILES for all instances of one pair member into the SMILES for the other pair member, according to rules specified by the user. The structural transformation rules are encoded by SMARTS; hence, the structural transformation can be chemically well defined.[26] When identifying matched pairs such as **3** and **4** that differ only by the addition of chlorine to an aromatic ring, we take each occurrence of ArCl and delete the chlorine (see Scheme 1). By matching the processed SMILES (appropriately standardized)[27] with the original SMILES set (also standardized), pairs of compounds that differ only by the desired change, such as those shown in Figure 2, can be found. The alternative method, implemented in the in-house program find_pairs,[28] requires the substructure defining each pair member to be coded in SMARTS and identifies molecules that are the same when these SMARTS hits are excluded. The connectivity of the atoms outside the SMARTS groups to those within must be mapped between the two. This is illustrated for the matched pairs such as **3** and **4** in the lower part of Scheme 1. These two different approaches are most appropriate to different kinds of structural transformation. There are numerous alternative ways of locating matched molecular pairs that will be more or less suitable for different kinds of structural transformation and to more or less global analyses.

Once the pairs of compounds have been identified, the corresponding data can be added and the difference or ratio computed, as appropriate, along with the appropriate statistics. In our case, this is carried out by the in-house program otsboho.[29] Solubility values are determined from agitation of compounds in 0.1 M phosphate buffer at pH 7.4 for 24 h at 25 °C. The supernatant is separated from undissolved material by double centrifugation and subsequently analyzed and quantified against a standard of known concentration in DMSO using generic HPLC−UV methodology coupled with mass spectral peak identification. Any measurement that is out of range of the assay is excluded from the analysis, and the lowest value of the measured solubilities is found for each compound that has been tested more than once. Variations in this

assay occur because the samples are often in different solid states: either polymorphs or amorphous. The lowest solubility is our best estimate of the solubility of the most stable polymorph. Protein binding is determined by equilibrium dialysis. A 20 μM concentration of compound is dialyzed against 10% plasma at a temperature of 37 °C for 18 h. The resulting samples are analyzed using generic HPLC−UV methodology coupled with mass spectral peak identification.[30] The reported K_1 value is the first apparent association constant [protein•ligand]/([protein][ligand]), all concentrations being measured in moles/liter. Any measurement that is out of the range of the assay is excluded from the analysis and the geometric mean of the first apparent association constant K_1 is computed for all compounds that have been tested more than once. Variations in this assay should not depend on the solid form, so the mean is taken to minimize changes due to experimental variation. The rat oral exposure assay involves the cocktail dosing of five test compounds plus a quality control compound (2 mg/kg per compound) to two rats. Plasma samples, which are obtained for up to 6 h after dosing, are analyzed by LC−MS−MS. The area under the curve (AUC) in h μg/mL is divided by the dose in mg/kg and the geometric mean value taken for those compounds that have been tested more than once. This is done in order to minimize effects due to experimental variation.

Supporting Information Available: Data concerning the interplane angle change in biphenyl compounds when *o*-chlorine is added, matched pairs analysis of human plasma protein binding, data on the conformational change experienced on going from compound **5** to **6**, pairwise comparison data (all three properties) for adding F and Cl to aromatic rings, and variation with lipophilicity of all three properties for the methylation of assorted heteroatoms. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Solubility models: (a) Delaney, J. S. Predicting aqueous solubility from structure. *Drug Discovery Today* **2005**, *10*, 289−295 and references therein. (b) Bergström, C. A. S.; Norinder, U.; Luthman, K.; Artursson, P. Experimental and computational screening models for prediction of aqueous drug solubility. *Pharm. Res.* **2002**, *19*, 182−188. (c) Huuskonen, J.; Salo, M.; Taskinen, J. Aqueous solubility prediction of drugs based on molecular topology and neural network modeling. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 450−456. (d) Xia, X.; Maliski, E.; Cheetham, J.; Poppe, L. Solubility prediction by recursive partitioning. *Pharm. Res.* **2003**, *20*, 1634−1640. (e) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **2001**, *46*, 3−26.

(2) Protein binding models: (a) Kratochwil, N. A.; Huber, W.; Müller, F.; Kansy, M.; Gerber, P. R. Predicting plasma protein binding of drugs−Revisited. *Curr. Opin. Drug Discovery Dev.* **2004**, *7*, 507−512 and references therein. (b) Kratchowil, N. A.; Huber, W.; Müller, F.; Kansy, M.; Gerber, P. R. Predicting plasma protein binding of drugs: A new approach. *Biochem. Pharmacol.* **2002**, *64*, 1355−1374. (c) Yamazaki, K.; Kanaoka, M. Computational prediction of the plasma protein-binding percent of diverse pharmaceutical compounds. *J. Pharm. Sci.* **2004**, *93*, 1480−1494. (d) Liu, J.; Yang, L.; Li, Y.; Pan, D.; Hopfinger, A. J. Prediction of plasma protein binding of drugs using Kier-Hall valence connectivity indices and 4D-fingerprint molecular similarity analyses. *J. Comput.-Aid. Mol. Design* **2005**, *19*, 567−583.

(3) Bioavailability models: (a) Martin, Y. C. A bioavailability score. *J. Med. Chem.* **2005**, *48*, 3164−3170. (b) Lu, J. J.; Crimin, K.; Goodwin, J. T.; Crivori, P.; Orrenius, C.; Xing, L.; Tandler, P. J.; Vidmar, T. J.; Amore, B. M.; Wilson, A. G. E.; Stouten, P. F. W.; Burton, P. S. Influence of molecular flexibility and polar surface area metrics on oral bioavailability in the rat. *J. Med. Chem.* **2004**, *47*, 6104−6107. (c) Jones, R.; Connolly, P. C.; Klamt, A.; Diedenhofen, M. Use of surface charges from DFT calculations to predict intestinal absorption. *J. Chem. Inf. Comput. Sci.* **2005**, *45*, 1337−1342.

(4) Yalkowsky, S. H.; Valvani, S. C. Solubility and partitioning I: Solubility of nonelectrolytes in water. *J. Pharm. Sci.* **1980**, *69*, 912−922.

(5) For instance (a) Klamt, A.; Eckert, F.; Hornig, M.; Beck, M. E.; Bürger, T. Prediction of aqueous solubility of drugs and pesticides with COSMO−RS. *J. Comput. Chem.* **2002**, *23*, 275−281. (b) Stukalin, E. B.; Korobov, M. V.; Avramenko, N. V. Solvation free energies of the fullerenes C_{60} and C_{70} in the framework of polarizable continuum model. *J. Phys. Chem. B* **2003**, *107*, 9692−9700.

(6) For instance, each table in the following recent papers corresponds to a set of matched molecular pairs: (a) Chen, C.; Stearns, B.; Hu, T.; Anker, N.; Santini, A.; Arruda, J. M.; Campbell, B. T.; Datta, P.; Aiyar, J.; Munoz, B. Expedited SAR study of high-affinity ligands to the $\alpha 2\delta$ subunit of voltage-gated calcium channels: Generation of a focused library using a solution-phase S_N2Ar coupling methodology. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 746−749. (b) Geneste, H.; Backfish, G.; Braje, W.; Delzer, J.; Haupt, A.; Hutchins, C. W.; King, L. L.; Kling, A.; Teschendorf, H.-J.; Unger, L.; Wernet, W. Synthesis and SAR of highly potent and selective dopamine D_3-receptor antagonists: 1*H*-Pyrimidin-2-one derivatives. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 490−494. (c) Ballard, P.; Bradbury, R. H.; Harris, C. S.; Hennequin, L. F. A.; Hickinson, M.; Johnson, P. D.; Kettle, J. G.; Klinowska, T.; Leach, A. G.; Morgentin, R.; Pass, M.; Ogilvie, D. J.; Olivier, A.; Warin. N.; Williams, E. Inhibitors of epidermal growth factor receptor tyrosine kinase: Novel C-5 substituted anilinoquinazolines designed to target the ribose pocket. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 1633−1637.

(7) Free, S. M., Jr.; Wilson, J. W. A mathematical contribution to structure−activity studies. *J. Med. Chem.* **1964**, *7*, 395−399.

(8) Kenny, P. W.; Sadowski, J. Structure modification in chemical databases. In *Chemoinformatics in Drug Discovery*; Oprea, T. I., Ed.; Wiley-VCH: Weinheim, 2004; 271−285.

(9) Sheridan, R. P.; Hunt, P.; Culberson, J. C. Molecular transformations as a way of finding and exploiting consistent local QSAR. *J. Chem. Inf. Comput. Sci.* **2006**, *46*, 180−192.

(10) Toogood, P. L.; Harvey, P. J.; Repine, J. T.; Sheehan, D. J.; VanderWel, S. N.; Zhou, H.; Keller, P. R.; McNamara, D. J.; Sherry, D.; Zhu, T.; Brodfuehrer, J.; Choi, C.; Barvian, M. R.; Fry, D. W. Discovery of a potent and selective inhibitor of cyclin-dependent kinase 4/6. *J. Med. Chem.* **2005**, *48*, 2388−2406.

(11) Yalkowsky, S. H.; He, Y. *Handbook of Aqueous Solubility Data*; CRC Press: Boca Raton, FL, 2003.

(12) (a) Allen, F. H. The Cambridge Structural Database: A quarter of a million crystal structures and rising. *Acta Crystallogr.* **2002**, *B58*, 380−388. (b) Bruno, I. J.; Cole, J. C.; Edgington, P. R.; Kessler, M.; Macrae, C. F.; McCabe, P.; Pearson, J.; Taylor, R. New Software for searching the Cambridge Structural Database and visualizing crystal structures. *Acta Crystallogr.* **2002**, *B58*, 389−397.

(13) Amidon, G. E.; He, X.; Hageman, M. J. Phys. (Paris) icochemical Characterization and Principles of Oral Dosage Form Selection. In *Burger's Medicinal Chemistry and Drug Discovery*, 6; Abraham, D. J., Ed.; John Wiley & Sons: New York, 2003; pp 2, 649−682.

(14) http://www.fda.gov/cder/guidance/3618fnl.htm.

(15) Lombardo, F.; Obach, R. S.; Shalaeva, M. Y.; Gao, F. Prediction of human volume of distribution values for neutral and basic drugs. 2. Extended data set and leave-class-out statistics. *J. Med. Chem.* **2004**, *47*, 1242−1250.

(16) Kragh-Hansen, U. Molecular aspects of ligand binding to serum albumin. *Pharmacol. Rev.* **1981**, *33*, 17−53.

(17) Fournier, T.; Medjoubi-N, N.; Porquet, D. Alpha-1-acid glycoprotein. *Biochim. Biophys. Acta* **2000**, *1482*, 157−171.

(18) The binomial probability distribution is used to determine the 95% confidence interval in the ratio of the observed number of property increase cases to the total number of observations. Computed with the tool available at http://members.aol.com/johnp71/confint.html.

(19) (a) Chou, J. T.; Jurs, P. C. Computer-assisted computation of partition coefficients from molecular structures using fragment constants. *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 172−178. (b) *clogP*, version 4.71; Daylight Chemical Information Systems Inc.: Santa Fe, NM.

(20) Abraham, M. H.; Duce, P. P.; Prior, D. V.; Barratt, D. G.; Morris, J. J.; Taylor, P. Hydrogen bonding. Part 9. Solute proton donor and proton acceptor scales for use in drug design. J. *J. Chem. Soc., Perkin Trans. 2* **1989**, 1355−1375.

(21) S_W is the solubility in water, P is the octanol−water partition coefficient, and mp is the melting point.

(22) (a) Obach, R. S.; Baxter, J. G.; Liston, T. E.; Silber, B. M.; Jones, B. C.; MacIntyre, F.; Rance, D. J.; Wastall, P. The prediction of human pharmacokinetic parameters from preclinical and in vitro metabolism data. *J. Pharmacol. Exp. Ther.* **1997**, *283*, 46−58. (b) Acharya, M. R.; Sparreboom, A.; Sausville, E. A.; Conley, B. A.; Doroshow, J. H.; Venitz, J.; Figg, W. D. Interspecies differences in plasma protein binding of MS-275, a novel histone deacetylase inhibitor. *Cancer Chemother. Pharmacol.* **2006**, *57*, 275−281.

(23) (a) Becke, A. D. A new mixing of Hartree−Fock and local-density-functional theories. *J. Chem. Phys.* **1993**, *98*, 1372−1377. (b) Becke, A. D. Density-functional thermochemistry. III. The role of exact

exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652. (c) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Chem. Phys.* **1994**, *98*, 11623−11627. (d) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle−Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785−789.

(24) Calculations were performed in *Gaussian 03*, revision C.02. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc.: Wallingford, CT, 2004.

(25) Navarro, E.; Alemán, C.; Puiggalí, J. Crystallographic and quantum mechanical results on Ψ[NHCO] aliphatic diamides. The number of methylenes strongly influences their structural and conformational properties. *Macromolecules* **1998**, *31*, 408−416.

(26) (a) *SMARTS*; Daylight Chemical Information Systems Inc.: Santa Fe, NM, Vol. 471. (b) http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html.

(27) Weininger, D.; Weininger, A.; Weininger, J. L. SMILES. 2. Algorithm for generation of unique SMILES notation. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 97−101.

(28) Jones, H. D.; Cosgrove, D. A. *find_pairs*; AstraZeneca Pharmaceuticals.

(29) Cosgrove, D. A. *otsboho*; AstraZeneca Pharmaceuticals.

(30) Wan, H.; Rehngren, M. High-throughput screening of protein binding by equilibrium dialysis combined with liquid chromatography and mass spectrometry. *J. Chromatogr. A* **2006**, *1102*, 125−134.

JM0605233