Journal of Medicinal Chemistry 56(6):2478-2486 (2013)
DOI:10.1021/jm301851v

How fragments are optimized? A retrospective analysis of 145 fragment optimizations.

György G. Ferenczy[a]* and György M. Keserű[b]*

[a]MTA-SE Molecular Biophysics Research Group, Semmelweis University, Tűzoltó u. 37-47, H-1094 Budapest, Hungary
[b]Discovery Chemistry, Gedeon Richter plc., Gyömrői út 19-21, H-1103 Budapest, Hungary; Present address: Research Centre for Natural Sciences, Hungarian Academy of Sciences, 1525 Budapest, P.O. Box 17, Hungary

Abstract

Fragment optimizations in nearly 150 fragment-based drug discovery programs reported in the literature during the last fifteen years were investigated. Analyzing physico-chemical properties and ligand efficiency indices we found that biochemical detection methods yield hits with superior ligand efficiency and lipophilicity indices than do X-ray and NMR. These advantageous properties are partially preserved in the optimization since higher affinity starting points allow optimizations better balanced between affinity and physico-chemical property improvements. Size-independent ligand efficiency (SILE) and lipophilic indices (primarily LELP) were found to be appropriate metrics to monitor optimizations. Small and medium enterprises (SME) produce optimized compounds with better properties than do big pharma companies and universities. It appears that the use of target structural information is a major reason behind this finding. Structure-based optimization was also found to dominate successful fragment optimizations that result in clinical candidates. These observations provide optimization guidelines for fragment-based drug discovery programs.

*Corresponding authors: Phone: +36 20 666 0232. Fax +36 1 266 6656. E-mail: ferenczy.gyorgy@med.semmelweis-univ.hu, (GGF); Phone: +36 1 438 1155, Fax: +36 1 438 1143, E-mail: gy.keseru@ttk.mta.hu (GMK)

Introduction

Fragment based drug discovery (FBDD) became a general approach used by both academic and industry players on a wide variety of targets in the last decade[1]. The foundation of FBDD was provided by two fundamental concepts, the theory of chemical space and molecular complexity. It has been suggested and later shown that the chemical space of low molecular weight compounds is much smaller than that of the more complex molecules. This results that the fragment space could be more effectively sampled by smaller collection of fragments relative to the conventional compound libraries screened in HTS campaigns. Based on their complexity theory Hann et al.[2] concluded that less complex fragments have typically higher chance forming favorable interactions with the binding site as compared to more complex compounds. Screening a limited set of low complexity compounds (fragments) would therefore improve the odds of identifying suitable chemical starting points for even the most challenging targets. Low molecular weight low complexity fragments, however, show limited affinity towards these binding sites. There are multiple conditions to detect the binding of fragments. Theoretically fragments should form high quality polar interactions within the binding site that suggests their low lipophilicity. At the technical level the identification of fragment hits requires high screening concentration that needs highly soluble hydrophilic fragments. Fragment screening therefore yields low complexity polar hits that were previously identified as preferred starting points for medicinal chemistry programs. These compounds could provide more operational freedom for structural modifications during multidimensional optimizations. On the other hand, it was concluded that due to their limited size and high polarity the physicochemical profile could be controlled more effectively in fragment optimizations. Limited affinity of the initial fragments, however, is typically far

from the general potency criteria of viable leads. Early optimization of fragments is therefore supposed to be potency driven that was recently identified as the major source of property inflation and molecular obesity[9]. This questions the highly controlled nature of fragment optimizations and in our view makes the process rather challenging. Considering the objectives of lead discovery teams they should optimize low millimolar or high micromolar affinity, low complexity polar fragments to low micromolar leads with promising specificity and acceptable physicochemical and ADME profile. Improvements in potency typically require increasing size and lipophilicity that have ambiguous effect on ADME properties. Although the permeability improves, metabolic stability is typically decreasing while promiscuity, P-gp and hERG liability are increasing with increasing lipophilicity. Due to the complex relationship between the potency and physicochemical and ADME properties and also the time pressure under these teams are working they need powerful tools delivering optimized compounds of high quality. Ligand efficiency metrics are of the primary choice and used guiding optimizations effectively. Ligand efficiency (LE) and its derivatives (BEI, SILE) are useful to control the size while lipophilic efficiency metrics (LLE, LLE_{Astex} and LELP) are used controlling the lipophilicity of the compounds optimized.

LE	$-RT\ln(K_d \text{ or } K_i)/N_{hev}$
BEI	$(pK_i \text{ or } pK_d)/MW$
SILE	$pIC_{50}/(N_{hev})^{0.3}$
LLE	pK_i - LogP (or LogD)
LLE_{Astex}	$0.11*\ln(10)*RT\{logP-\log(K_d \text{ or } K_i \text{ or } IC_{50})\}/N_{hev}$
LELP	logP/LE

The pioneering work of Reynolds and coworkers revealed that ligand efficiency is size dependent. Since the size of the fragments increases significantly during their optimization the use of size independent metrics such as SILE would be more meaningful in fragment settings. Similarly, lipophilic efficiency could be more straightforwardly evaluated by LELP and LLE_{Astex} that also reflect the molecular size. Validation of these assumptions was also in the scope of this study that was originally initiated to answer the following questions: i) do fragment starting points eliminate the risk of property inflation per se, ii) how do ligand efficiency metrics support fragment optimizations, and iii) what is the impact of the detection methods, the hit properties, the optimization strategy and the company culture on the outcome of the optimizations. To achieve these goals we collected and analyzed fragment optimizations published up to 2011 and report the results of this investigation here.

Methods

A literature search of fragment hit optimizations up to and including year 2011 has been performed. 145 fragment programs were identified from major reviews[1,9,] and further examples published between 2009 and 2011 (see Supplementary Information). They targeted 83 proteins out of which 6 are receptors, one is an ion channel and 76 are enzymes. A database was set up with pairs of fragment hits and their optimized counterparts containing their chemical structures, affinities and calculated properties as obtained with the calculator plugins from ChemAxon. The physico-chemical and ligand efficiency parameters were analyzed. It is to be noted that the level of optimization is varying and optimized compounds include leads, tools and clinical candidates. This may contribute to the observed spread of properties. Distributions of the parameters for the hits and for the optimized fragments are

presented in histograms generated with Microsoft Excel. The analysis is primarily based on the median of the distributions. Differences between properties were investigated by calculating the Mann-Whitney U test with Statistica and was accepted to be significant at $p<0.05$, unless otherwise noted.

Results and Discussions

Physico-chemical and Ligand Efficiency Parameters

Fragments are small sized compounds usually with low affinity towards their targets and thus affinity improvement is a fundamental objective of their optimizations. Indeed, the median of the affinity of the fragment hits is 80 μM and that of the optimized compounds is 60 nM representing an approximately 3 order of magnitude improvement (Median of the affinity before and after optimization together with the medians of other properties discussed below are shown in Table 1.) The affinity improvement exceeds 1 order of magnitude in 90% of the cases (and exceeds 1.5 in 80%). The affinity distributions are shown in Figure 1. In the course of affinity improvement the size and the polarity of the fragments change significantly. The histograms of the number of heavy atoms for the fragments and for the optimized compounds together with the histogram of their changes are shown in Figure 2. The median of the molecular weight of the fragment hits is 215 Da with a heavy atom number of 15. The size increases considerably in the optimizations. 90% of the optimizations is accompanied by an increase of at least 4 heavy atoms corresponding to an increase of 50 Da in molecular weight. The median of the number of heavy atoms of the optimized compounds is 28 and that of the molecular weights is 381. Lipophilicity increases from the initial logP of 1.79 to 3.19 (see

Figure 3). 80% of the optimizations result in compounds with a logP higher than the corresponding starting fragment.

Fragment hits have typically low affinity but owing to their small size they have reasonable ligand efficiency. The median of the ligand efficiency for the fragments is 0.37 and it hardly changes during the optimizations as it is shown by the median of 0.34 of the optimized compounds and can also be seen in Figure 4. Maximal ligand efficiency, however, is known to be size dependent[7], and fragment optimizations are accompanied with significant increase in molecular size and complexity (*cf.* Figure 2). This means that typically an increase in affinity at constant LE is achieved with an increased molecular size (*cf.* with LE definition). This is an advantageous step forward, since maximal LE decreases with molecular size, and thus a compound with increased size and unaltered LE is a further step ahead to the maximal LE. This improvement is quantitatively expressed by size independent ligand efficiency (SILE). Indeed, SILE increases from 1.75 to 2.64 in the optimizations and this significant increase is also apparent from the histograms in Figure 5. This finding emphasizes the importance of taking into account the effect of the compound size on the ligand efficiency in order to obtain a sensible measure of compound quality changes in optimizations. Ligand-lipophilicity efficiency increases during the optimizations from 2.15 to 4.13 (Figure 6). Even the optimized compounds have lower LLE than the assumed ideal value of 5. Nevertheless, the nearly 2 order of magnitude increase in ligand-lipophilicity efficiency and the over 3 order of magnitude increase in affinity (see above) are in line with the modest increase of logP that was found to be less than 1.5 units. The LELP parameter that has been shown to relate to the ADME properties of compounds has a favorable median of 4.62 for the fragment hits that increases to the less favorable value of 8.47 for the optimized compounds (Figure 7). This unfavorable tendency is associated with a fairly wide range of LELP values for the optimized

compounds suggesting that various optimization strategies lead to compounds with considerably different quality.

A comparison of the efficiency and size change as observed in optimizations is shown in Figure 8. These results show similarity to the analogous results obtained for “successful drug discovery programs” (Figure 5 in ref), except that the size increase is more pronounced and no lowering of MW is observed in our analysis that is a logical consequence of considering fragment optimizations. Ligand efficiency is decreased in almost 60% of optimizations that appears to be similar to the value reported in ref. . By contrast, when the size dependence of the ligand efficiency is taken into account by depicting MW ratio against SILE ratio in Figure 9, then SILE is improved in 90% of the optimizations. This finding again points out the importance of using a size dependent measure for characterizing the efficiency of compounds with different size. This is particularly relevant in fragment optimizations that are accompanied with a significant increase in ligand size.

Finally we analyzed changes in physicochemical properties and ligand efficiency metrics in the context of medicinal chemistry optimizations (Table 2). Classical HTS based optimizations starts from hits with micromolar affinity that is typically improved by 1.4 pPotency unit to provide nanomolar compounds. This change in potency is associated with moderate increase in both molecular weight and lipophilicity. Ligand efficiency does not change very much, however, SILE increases due to the 3-4 non-hydrogen atoms added. Ligand-lipophylic efficiency (LLE) improves, but LELP is virtually constant. In contrast, the more improvement in potency needed for the optimization of less potent fragment hits results in much higher increase in both molecular weight and logP. Although the average increase in heavy atoms is about 13-14, SILE improves significantly but this is not reflected in LE that seems to be constant again. Improvement in LLE is comparable to that observed for HTS based optimizations, however, in line with the 4-fold higher increase in lipophilicity the size

dependent LELP is getting significantly worse. Comparing these changes to fragment optimization resulting clinical candidates one can conclude that in successful optimizations the somewhat higher increase in potency is achieved by smaller increase in size and much smaller in lipophilicity. Interestingly, LE does not discriminate these processes but SILE and LLE improves more significantly than average and the undesired change in LELP is also much smaller. These trends are pretty much similar to those reported for successful lead optimizations demonstrating effective control in both size and lipophilicity as indicated by large improvements in SILE, LLE and even in LELP. Based on this analysis we argue that similar to lead optimizations the strict control of physicochemical properties using SILE, LLE and LELP would help very much identifying high quality compounds from fragment optimizations.

The Impact of Detection Methods on the Properties of Fragment Hits and Optimized Compounds

Next we investigated the effect of hit discovery strategies on the quality of initial and optimized hits. Four methods for detecting hits are well represented in our dataset of fragment optimizations, namely biochemical (38%), X-ray (18%), NMR (25%) and virtual screening (11%). Other methods, like SPR and MS represent only 8% and they were not included in the analysis below. It should be noted that hit identification often includes more than a single method (e.g. biochemical+NMR, or virtual+biochemical) and in these cases the first applied method was selected. Virtual screening is unique in the sense that its hits are always validated by another method, most often by biochemical tests. Another characteristic feature of virtual

screening is that it typically exploits preliminary structural information either on ligands or on the target and eventually on both.

A comparison of the parameters of the hits obtained with the four methods shows significant differences in five pairs: virtual-NMR, biochemial-NMR, virtual-X-ray, NMR-X-ray and biochemical-X-ray. Hit parameters are more advantageous for the virtual and biochemical methods than for X-ray and NMR the differences being more pronounced for the latter (Table 3). In the following detailed analysis property values are quoted where significant differences observed.

Biochemical (4.75) and virtual (4.73) hits have higher affinity than do X-ray (3.56) and NMR (3.53) hits. Although X-ray hits have the smallest size among all the methods (HA=13,0 vs. biochemical 15.5, NMR 15.0, virtual 17.0) SILE is higher for biochemical (2.19) than for X-ray (1.68). The higher affinity and the similar size yield more beneficial SILE for virtual (2.02) and biochemical (2.19) than for NMR (1.53) screens. The difference in SILE between biochemical (2.19) and X-ray (1.68) hits is the consequence of the higher affinity of biochemical hits that overbalances the smaller size of the X-ray hits. Significant differences in ligand efficiency were found for the biochemical (0.40) vs. NMR (0.31) and for X-ray (0.38) vs. NMR (0.31). LLE is more advantageous for biochemical (3.10) and virtual (3.00) than for X-ray (1.83) and NMR (1.36). ($p=0.070$ for virtual vs. X-ray). LELP exhibits difference in favor of biochemical over NMR (4.05 vs. 5.65).

The above analysis points out that hit properties exhibit significant dependence on hit detection methods. This can be partially traced back to the difference in the compound libraries of the hit detection methods. Two principal motifs behind library compilation are the objective of the screening (e.g. fragments vs. larger compounds) and the technical constraints of the hit detection method (e.g. moderate potency for NMR or small library size for X-ray).

Table 3 also shows how the properties of optimized compounds depend on the hit detection method. Significant differences are observed only in few cases for the optimized compounds suggesting that the optimization tends to diminish the differences in hit properties. Interestingly, however, hits identified by biochemical methods preserve their advantage after optimization (see Table 3) primarily relative to NMR hits (pIC_{50} 7.68 vs. 6.70, LE 0.37 vs. 0.32, LLE 4.55 vs. 3.17, LELP 7.62 vs. 10.51, SILE 2.91 vs. 2.45), but also relative to X-ray hits (pIC_{50} 7.68 vs. 7.13 with p=0.055, SILE 2.91 vs. 2.57).

The effect of hit quality on the optimization process and on the optimized compound was further investigated by using data from 18 optimizations where one or several intermediates in the optimization are also revealed and their activities are available. Figure 10 plots the activity vs. molecular weight for these optimizations. Best fit linear trends are also shown. These lines contain information on the effectiveness of the optimization; the slope is smaller when 1 unit affinity improvement is achieved with a smaller increase in molecular weight thus a smaller slope corresponds to a more efficient optimization. The slope varies between 10.3 and 138.5 with an average of 61.4 and a standard deviation of 34.3. The average is remarkably similar to that (64) reported in ref , where a pK_d vs. MW plot obtained by an *a posteriori* deconstruction of lead compounds. The variation of the slope, however, is more important in our analysis. This suggests that real optimizations do not necessarily follow the ideal path proposed on the basis of *a posteriori* lead deconstruction. The assumption of the ideal optimization path is equivalent with assuming that hit properties (affinity and molecular weight) well determine the quality of optimized compounds. Our analysis based on true optimizations gives a more complex picture. Data in Table 3 and in Figure 10 show that significant differences observed at hit level may disappear in optimized compounds. On the other hand, biochemical hits appear to be exceptions as these hits tend to result in optimized compounds superior to those optimized from hits with less advantageous properties. We propose the following explanation

for the above findings. The affinity driven optimization of the generally low affinity hits is a common practice in the hit-to-lead process of fragment based optimizations. This practice tends to diminish the differences originating from the hit detection method. The apparent exceptions are biochemical hits that have higher affinities and therefore, a more balanced optimization with less aggressive affinity improvement is possible. This explains why biochemical hits are able to preserve their advantageous properties in the optimizations and why they yield optimized compounds with better affinity and lipophilicity indices with respect to those optimized from NMR hits and, to a lesser extent, from X-ray hits. In the latter case we think it is the use of structural information that helps improving the affinity without jeopardizing the physicochemical profile.

The effect of hit properties to optimized compounds will be further investigated in a comparative analysis of all and successful optimizations the latter resulting in clinical candidates (see later).

Company Culture and Fragment Optimization

Optimizations in drug discovery programs have been shown to be affected by company culture. In order to investigate this point, changes in compound parameters in fragment-based optimizations performed by academic (UNI 18%), small and medium enterprises (SME, 37%) and big pharmaceutical companies (BIG, 45%) were compared. In the following analysis values of significant property changes are quoted.

Interestingly, data show that SME results are superior to BIG and UNI results. Affinity improvements in SME are larger than in BIG (3.10 vs. 2.35) and this appears also in SILE changes (0.84 vs. 0.60). Considering the increase of LLE SME (1.84) perform better than BIG

(0.88) and UNI (0.70). These data show that better affinity improvements in SME are not compensated by a better control of other parameters (size, logP) by BIG and UNI. We propose that the reason behind the good performance of SME is the dominant application of protein structural information in optimizations and the company culture that emphasizes its utmost importance. SME are typically platform-companies whose proprietary technology often includes biophysical methods that provide different levels of structural information. Indeed, it was found that 40 out of the 53 optimizations performed by SME used protein structural information (75%). By contrast, this ratio is 41/65 (63%) for BIG. The lower ratio is probably the consequence of the availability of a wide range of technologies whose use is traditionally part of the optimization process. Nevertheless we could not rule out that the higher content of challenging targets without structural information in big pharma portfolios might also contribute to this observation. This is indicated by the fact that BIG are less restrictive in their target selection: all but one programs (six out seven) with non-enzyme targets are associated with BIG. The use of structural information is much less frequent by UNI (33%), most probably because of the high entry level investment structure-based optimizations require. Considering the optimizations of all the 145 fragment-based programs analyzed here 15 (10%) and 75 (52%) of them used structural information from NMR and X-ray crystallography, respectively. Thus 62% of the optimizations used structural information and it is likely to exceed the use of structural information in drug discovery programs in general. This tendency goes parallel to recent polls published at the Practical Fragments blog. The apparent consensus on the importance of atomic level information on the target structure in fragment-based programs is amplified by the present finding that the better performance of SME is accompanied by a more frequent use of structure-based optimization.

Clinical Candidates from Fragment Optimizations

Up to now 18 clinical candidates were published as the results of fragment-based drug discovery programs, (see also Supporting Information). The structure of the starting point and the clinical candidate was identified for 13 programs. This represents approximately 10 % of our dataset. These optimizations were compared to all fragment-based optimizations. The limitations of this comparison are the small number of clinical candidates with available structures and also the varying level of optimization of the compounds taken from the literature.

Properties of programs leading to clinical candidates are indicated in Figure 1- Figure 7 together with those of all fragment-based programs. These figures show that clinical candidates originated from hits with similar size and lipophilicity as hits of other programs. On the other hand, their affinity is higher: pXC50=7.85 vs. 7.22, but interestingly it is not higher than that found for average oral drugs. This suggests that compounds optimized to the affinity of pXC50~8 have almost the optimal combination of potency and physico-chemical/ADME properties maximizing their in vivo efficacy. Owing to their higher affinity accompanied with a size and lipohilicity similar to average optimizations the size-independent ligand efficiency, SILE (2.89 vs. 2.64), and ligand-lipophilicity efficiency, LLE (4.91 vs. 4.13 p=0.058) of compounds from successful fragment optimizations are also significantly better. In the course of optimizing their ligand efficiency, LE hardly changes, SILE improves and logP increases modestly, and thus LLE (ΔLLE=3.17 vs. 1.49) and LELP (ΔLELP=1.38 vs. 4.02) exhibit significantly more beneficial changes than do average optimizations. These observations clearly show that lipophilic efficiency monitoring, whose importance is already established for drug discovery programs in general, has fundamental importance in fragment-based optimizations, as well.

It is worth also noting that 10 out of 13 clinical candidates come from optimizations that used target structural information. This high representation of structure-based optimizations in successful fragment projects underlines the importance of exploiting atomic level structural information on the target.

The properties of all compounds optimized from fragment hits were compared to the properties of those compounds that reached clinic. This was done with the objective to see if certain properties are appropriate to indicate the success of optimizations. The most striking observation is the concentration of the clinical candidates around the preferred region in LELP vs. logP diagram (Figure 11). It was previously established that LELP under 10 and logP under 5 tend to be associated with beneficial physico-chemical and ADME properties. The appearance of the clinical candidates exclusively within or near to this preferred region while the wider spread of other optimized compounds in the LELP-logP space shows, that these are useful parameters to be considered for fragment optimization programs. It is instructive to see, that the separation of clinical candidates from other optimized compounds is more pronounced in the LELP space. In fact, there is no compound with logP>5 and LELP<10, while there are several compounds with LELP>10 and logP<5. Considering that LELP well discriminates preferred starting points of fragment optimizations the current finding underpins the use of LELP as a valuable parameter also in the optimization process.

Conclusion

The analysis of nearly 150 fragment optimizations published up to 2011 reveals several interesting features and trends. The reduced size of the fragment space, the advantageous physico-chemical properties and high ligand efficiencies of fragment hits are all in favor of

fragment-based programs. On the other hand, the typically low affinity of fragment hits makes their optimizations to leads and later to clinical candidates challenging. Ligand efficiency and lipophilicity indices are particularly important for fragment optimizations owing to the significant property changes inherent in the hit-to-lead-to-candidate process. Moreover, the considerable increase in molecular size makes the use of size dependent indices unpractical and even misleading. Fragment optimizations are accompanied with important affinity increase that is well reflected by increasing SILE values while LE, its size-dependent variant, does not change in a consistent manner. Considering lipophilicity metrics, logP and LELP were found to be particularly useful as all clinical candidates arose from fragment-based programs were located in, or near to the logP 0-5 and LELP 0-10 space.

Hit detection methods were found to affect hit quality and biochemical hits were found to be especially advantageous. Biochemical hits not only exhibit better affinity and lipophilicity indices over NMR and X-ray hits, but compounds optimized from them preserve their advantageous properties over those optimized primarily from NMR and partly from X-ray hits. We propose that this is due to their higher affinity that allows a less aggressive potency optimization and consequently optimized compounds with more balanced properties. In addition, we think that biochemical screening has a higher chance identifying functionally active, pharmacologically relevant hits. On the other hand, however, we underline that the generally high false positive rate of fragment screening urges orthogonal biophysical profiling at the hit validation phase.

In addition to the hit detection method, the way of optimization also affects the properties of the optimized compounds. It appears that the use of atomic level structural information of the target (most often from X-ray but also from NMR) advantageously influences compound properties. It was found that SME that apply almost exclusively structure-based optimizations are particularly successful in producing good quality optimized compounds. By contrast, big

pharma and academia that apply structure-based optimizations less frequently produce, on average, somewhat inferior compounds at least in the present set of fragment optimization programs. The high representation of SME (10/13) and that of structure-based optimizations (10/13) in fragment-based clinical candidates supports this view.

The results of the analysis allow us to make recommendations for fragment-based drug discovery programs. We propose hit-detection by biochemical methods, followed by orthogonal biophysical validation and subsequent structure-based optimization with a target affinity of pIC_{50}~8. Large scale analysis of fragment optimizations suggests that monitoring size-independent ligand efficiency (SILE) and lipophilic efficiency (LELP) indices increase the chance of success in fragment-based drug discovery programs.

Supporting Information Available: All hits and optimized compounds together with their parameters and references to the original papers. This material is available free of charge via the Internet at **http://pubs.acs.org**.

Abbreviations Used

BEI, Binding efficiency index; LELP, ligand-efficiency-dependent lipophilicity; LLE, Ligand-lipophilicity efficiency; SILE, Size independent ligand efficiency; SPR, Surface plasmon resonance

References

1. Congreve, M,.; Chessari, G,; Tisi, D.; Woodhead, A. J. Recent developments in fragment-based drug discovery. *J Med Chem.* **2008,** *51,* 3661-3680.

2. Murray, C. W.; Verdonk, M. L.; Rees, D. C. Experiences in fragment-based drug discovery. *Trends in Pharmacological Sciences* **2012,** *33,* 224-232.

3. Blum, L. C.; Reymond, J.-L. 970 Million Druglike Small Molecules for Virtual Screening in the Chemical Universe Database GDB-13. *J. Am. Chem. Soc.* **2009,** *131,* 8732-8733.

4. Bohacek, R. S.; McMartin, C.; Guida, W. C. The art and practice of structure-based drug design: A molecular modeling perspective. *Med. Res. Rev.* **1996**, *16*, 3–50.

5. Hann, M. M.; Leach, A. R.; Harper, G. Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 856–864.

6. Leach, A. R., Hann, M. M. Molecular complexity and fragment-based drug discovery: ten years on. *Curr. Op. Chem. Biol.* **2011,** *15,* 489-496.

7. Ferenczy, G. G.; Keserű, G. M. Thermodynamics of fragment binding. *J. Chem. Inf. Model.* **2012,** *52,* 1039-1045.

8. Teague, S.J.; Davis, A. M.; Leeson, P. D.; Oprea, T. The Design of Leadlike Combinatorial Libraries. *Angew. Chem. Intl. Ed. Eng.* **1999,** *38,* 3743-3747.

9. Murray, C.W.; Rees, D.C. The rise of fragment-based drug discovery. *Nature Chemistry* **2009,** *1,* 187-192.

10. Ferenczy, G. G.; Keserű, G. M. Thermodynamics guided lead discovery and optimization. *Drug. Discov. Today* **2010,** *15,* 919-932.

11. Keserű, G. M.; Makara, G. M.; The influence of lead discovery strategies on the properties of drug candidates. *Nat. Rev. Drug Discov.* **2009,** 8, 203-12.

12. Hann, M. M. Molecular obesity, potency and other addictions in drug discovery *Med. Chem. Commun.* **2011,** *2,* 349-355.

13. Gleeson, M. P.; Hersey, A.; Montanari, D.; Overington, Probing the links between in vitro potency, ADMET and physicochemical parameters. *Nat. Rev. Drug Discov.* **2011,** *10,* 197-208.

14. Waring, M. Lipophilicity in drug discovery. *Expert Opin. Drug Discov.* **2010,** *5,* 235–248 (2010).

15. Reynolds, C. H.; Bembenek, S. D.; Tounge, B. A. The role of molecular size in ligand efficiency. *Bioorg. Med. Chem. Lett.* **2007**, *17,* 4258–4261.

16. Alex, A. A.; Flocco, M. M. Fragment-based drug discovery: What has it Achieved so Far? *Curr. Top. Med. Chem.* **2007,** *7,* 1544-1567.

17. JChem for Excel 5.11.3.822, 2012, ChemAxon (http://www.chemaxon.com)

18. StatSoft Inc. (2010) STATISTICA (data analysis software system), version 10. www.statsoft.com.

19. Nissink, J. W. M. Simple Size-Independent Measure of Ligand Efficiency. *J. Chem. Inf. Model.* **2009**, *49*, 1617–1622.

20. Leeson, P. D.; Springthorpe B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nature Rev. Drug Discov.* **2007,** *6,* 881–890.

21. Perola, E. An Analysis of the Binding Efficiencies of Drugs and Their Leads in Successful Drug Discovery Programs. *J. Med. Chem.* **2010,** *53,* 2986–2997.

22. de Kloe, G. E.; Bailey, D.; Leurs, R.; de Esch, I. J. Transforming fragments into candidates: small becomes big in medicinal chemistry. *Drug Discov Today* **2009**, *14*, 630-46.

23. Hajduk, P. J. Fragment-Based Drug Design: How Big Is Too Big? *J. Med. Chem.* **2006,** *49,* 6972-6976.

24. Leeson, P. D.; St-Gallay, S. A. The influence of the 'organizational factor' on compound quality in drug discovery. *Nat. Rev. Drug. Discov.* **2011,** *10,* 749-765.

25. http://practicalfragments.blogspot.hu/2012/08/poll-results-do-you-need-structure.html

26. Erlanson, D. A. Introduction to Fragment-Based Drug Discovery *in Fragment-Based Drug Discovery and X-Ray Crystallography* (Eds: Thomas G. Davies, Marko Hyvönen), Topics in Current Chemistry, Volume 317, pp 1-32., Springer, New York, 2012.

27. Baker, M.: Fragment-based lead discovery grows up. *Nature Rev. Drug. Discov.* **2013,** *12,* 5-10.

28. Hann, M. M.; Keserű, G. M.: Finding the sweet spot:the role of nature and nurture in medicinal chemistry. *Nature Rev. Drug Discov.* **2012,** *11,* 355-365.

29. Tarcsay, Á.; Nyíri, K.; Keserű, G. M. Impact of Lipophilic Efficiency on Compound Quality. *J. Med. Chem.* **2012,** *55,* 1252–1260.

30. Wager, T. T.; Chandrasekaran, R. Y.; Hou, X; Troutman, M. D.; Verhoest, P. R.; Villalobos, A.; Will, Y. Defining desirable central nervous system drug space through

the alignment of molecular properties, in vitro ADME, and safety attributes. *ACS Chem. Neurosci.* **2010,** *1*, 420–434

31. 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 Del Rev.* **2001,** *46*, 3–26.

Table 1
Medians of properties for fragment hits and for optimized compounds

	fragment hit	optimized compound
pIC_{50}	4.10	7.22
HA	15	28
logP	1.79	3.19
LE	0.37	0.34
SILE	1.75	2.64
LLE	2.15	4.13
LELP	4.62	8.47

Table 2
Ligand efficiency metrics in fragment optimizations (mean of property changes)

Process	pPot change	MW change	logP change	LE change	SILE change	LLE change	LELP change
HTS based opt. from ref.	1.39	51.5	0.27	0.02	0.58	1.1	0.1
Fragment opt. from this work	2.74	186.9	1.33	-0.04	0.70	1.4	4.8
Fragment opt. (successful) from this work	3.10	165.5	0.48	-0.01	0.86	2.6	1.0
Lead opt. (successful) from ref.	2.08	89.9	0.05	0.01	0.85	2.1	-1.1

Table 3
Property medians for compounds grouped by hit detection methods. Values are shown for pair of detection methods where properties are not equal according to the Mann-Whitney U-test at a 0.05 significance level. The analysis is performed separately for hits and for optimized compounds.

		Bio-NMR	*Bio-Xra*	*NMR-Vir*	*NMR-Xra*	*Vir-Xra*
pIC50	*hit*	4.75-3.53	4.74-3.56	3.53-4.73		4.73-3.56
	opt	7.68-6.70	7.68-7.13[a]			
MW	*hit*		223-190			235-190
	opt					
HA	*hit*		15.5-13.0		15.0-13.0	17.0-13.0
	opt					
LE	*hit*	0.40-0.31			0.31-0.38	
	opt	0.37-0.32				
LLE	*hit*	3.10-1.36	3.10-1.83	1.36-3.00		3.00-1.83[b]
	opt	4.55-3.17				
LELP	*hit*	4.05-5.65				
	opt	7.62-10.51				
SILE	*hit*	2.19-1.53	2.19-1.68	1.53-2.02		
	opt	2.91-2.45	2.91-2.57			

[a] p=0.055
[b]p=0.070

Figure 1
a)
Affinity distribution of fragment hits (FRG) and optimized compounds (OPT). Vertical lines show median of hits (red) and optimized compounds (blue). Medians of hits and optimized compounds for programs resulting clinical candidates are shown under the horiziontal axis.
b)
Affinity change distribution calculated from the differences in affinities of optimized compounds and fragment hits.

Figure 2
a)
Number of heavy atom distribution of fragment hits (FRG) and optimized compounds (OPT). Vertical lines show median of hits (red) and optimized compounds (blue). Medians of hits and optimized compounds for programs resulting clinical candidates are shown under the horiziontal axis.
b)
Heavy atom change distribution calculated from the differences in the number of heavy atoms of optimized compounds and fragment hits.

Figure 3
a)
logP distribution of fragment hits (FRG) and optimized compounds (OPT). Vertical lines show median of hits (red) and optimized compounds (blue). Medians of hits and optimized compounds for programs resulting clinical candidates are shown under the horiziontal axis.
b)
logP change distribution calculated from the differences in logP of optimized compounds and fragment hits.

Figure 4
a)
Ligand efficiency (LE) distribution of fragment hits (FRG) and optimized compounds (OPT). **Vertical** lines show median of hits (red) and optimized compounds (blue). Medians of hits and optimized compounds for programs resulting clinical candidates are shown under the horiziontal axis. These latter are very close (0.39 and 0.40) although they belong to different bins.
b)
LE change distribution calculated from the differences in LE of optimized compounds and fragment hits.

Figure 5
a)
Size independent ligand efficiency (SILE) distribution of fragment hits (FRG) and optimized compounds (OPT). Vertical lines show median of hits (red) and optimized compounds (blue). Medians of hits and optimized compounds for programs resulting clinical candidates are shown under the horiziontal axis.
b)
SILE change distribution calculated from the differences in SILE of optimized compounds and fragment hits.

Figure 6
a)
Ligand-lipophilicity efficiency (LLE) distribution of fragment hits (FRG) and optimized compounds (OPT). Vertical lines show median of hits (red) and optimized compounds (blue). Medians of hits and optimized compounds for programs resulting clinical candidates are shown under the horiziontal axis.
b)
LLE change distribution calculated from the differences in LLE of optimized compounds and fragment hits.

Figure 7
a)
Lipophilic efficiency dependent lipophilicity (LELP) distribution of fragment hits (FRG) and optimized compounds (OPT). Vertical lines show median of hits (red) and optimized compounds (blue). Medians of hits and optimized compounds for programs resulting clinical candidates are shown under the horiziontal axis.
b)
LELP change distribution calculated from the differences in LELP of optimized compounds and fragment hits.

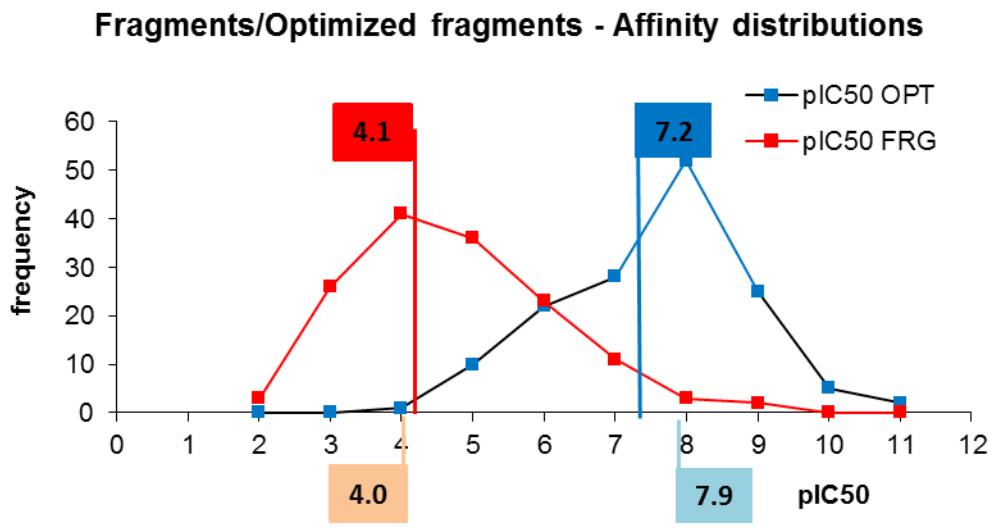
Figure 8
Ligand efficiency (LE) ratio vs. molecular weight (MW) ratio. LE ratio = LE(optimized)/LE(fragment) and MW ratio = MW(optimized)/MW(fragment).

Figure 9
Size independent ligand efficiency (SILE) ratio vs. molecular weight (MW) ratio. SILE ratio = SILE(optimized)/SILE(fragment) and MW ratio = MW(optimized)/MW(fragment).

Figure 10
Ligand affinity (pKi) vs. molecular weight (MW) for 18 fragment optimization programs. Best fit linear trends are also shown.

Figure 11
logP vs. LELP for compounds optimized from fragment hits. Green rectangles correspond to compounds that reached clinic. They are within or near to the green area of preferred LELP and logP values.

Figure 1
a)


Fragments/Optimized fragments - Affinity distributions
pIC50 OPT
pIC50 FRG
4.1
7.2
4.0
7.9
frequency
pIC50

b)

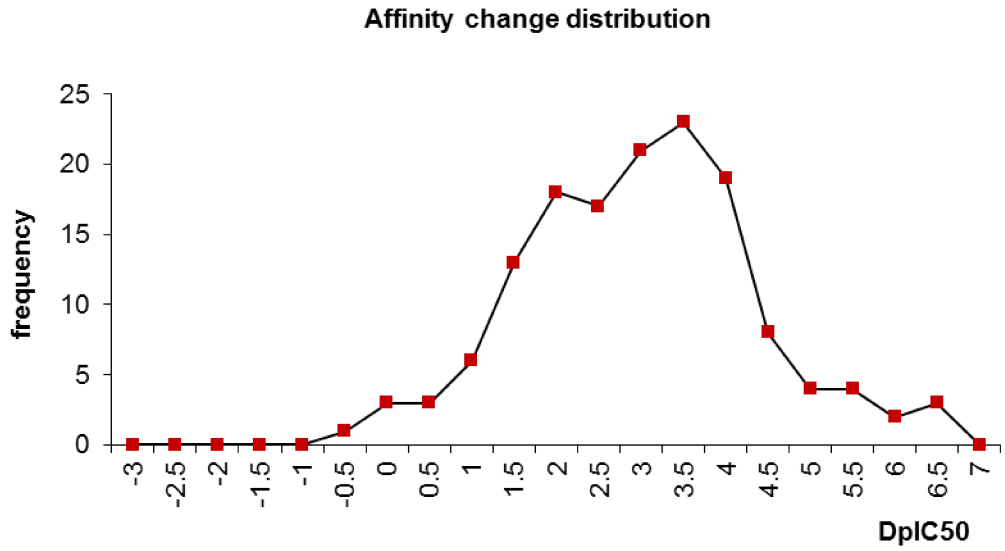
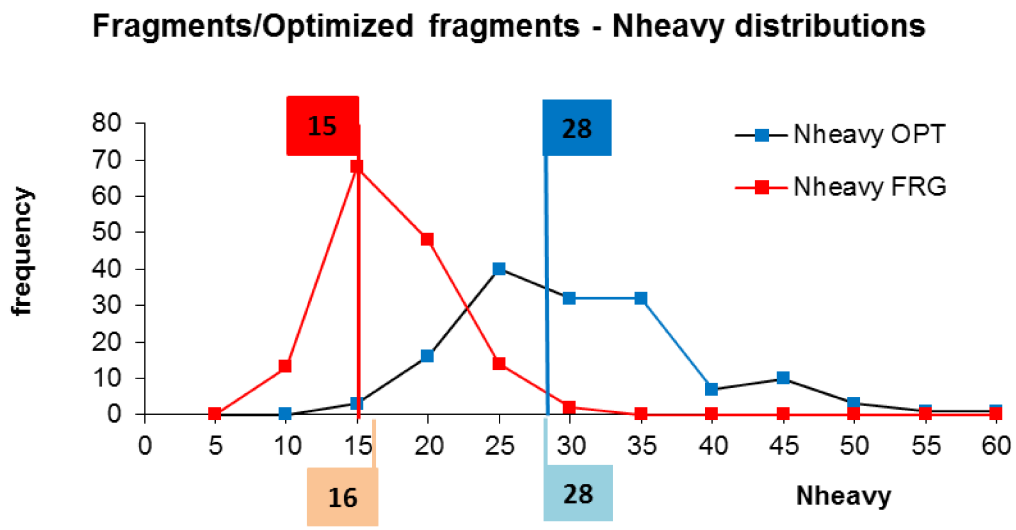

Affinity change distribution
frequency
DpIC50

Figure 2

a)


Fragments/Optimized fragments - Nheavy distributions
frequency
Nheavy
Nheavy OPT
Nheavy FRG
15
28
16
28
0
10
20
30
40
50
60
70
80
0
5
10
15
20
25
30
35
40
45
50
55
60

b)

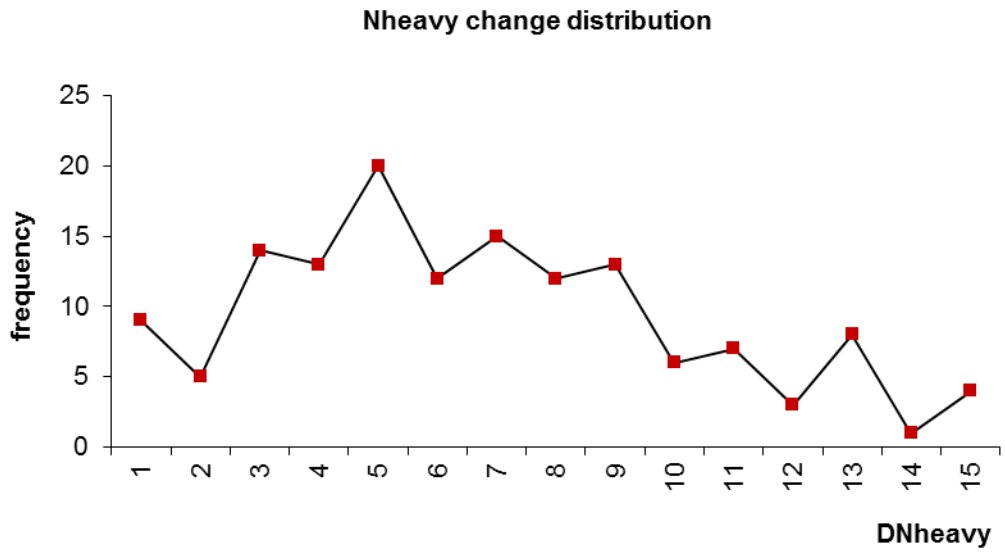
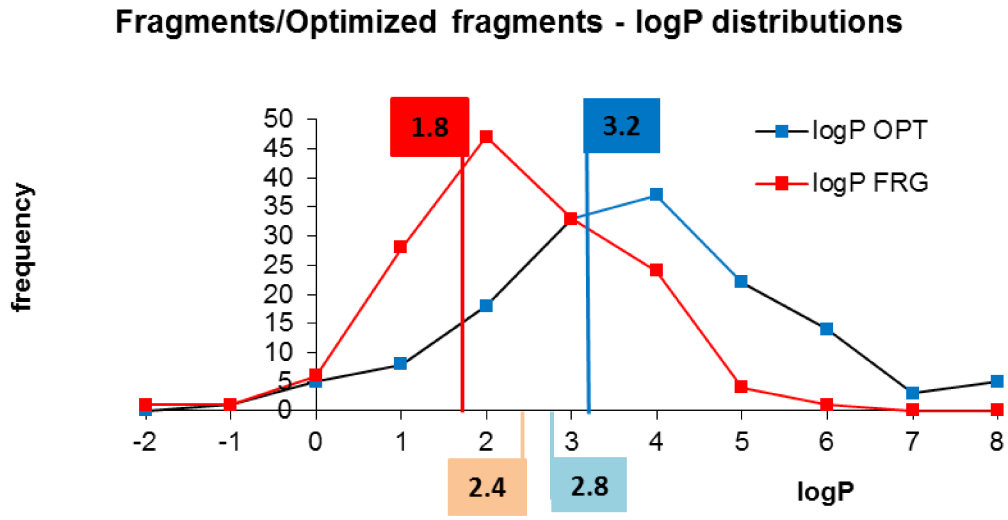

Nheavy change distribution
frequency
DNheavy
0
5
10
15
20
25
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

Figure 3

a)


Fragments/Optimized fragments - logP distributions
frequency
50
45
40
35
30
25
20
15
10
5
0
-2
-1
0
1
2
3
4
5
6
7
8
1.8
3.2
2.4
2.8
logP OPT
logP FRG
logP

b)

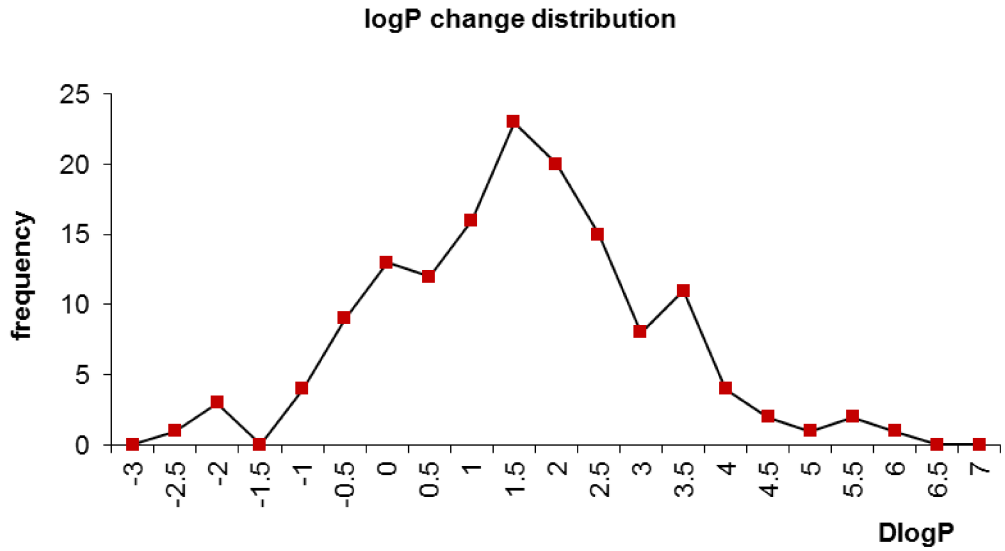
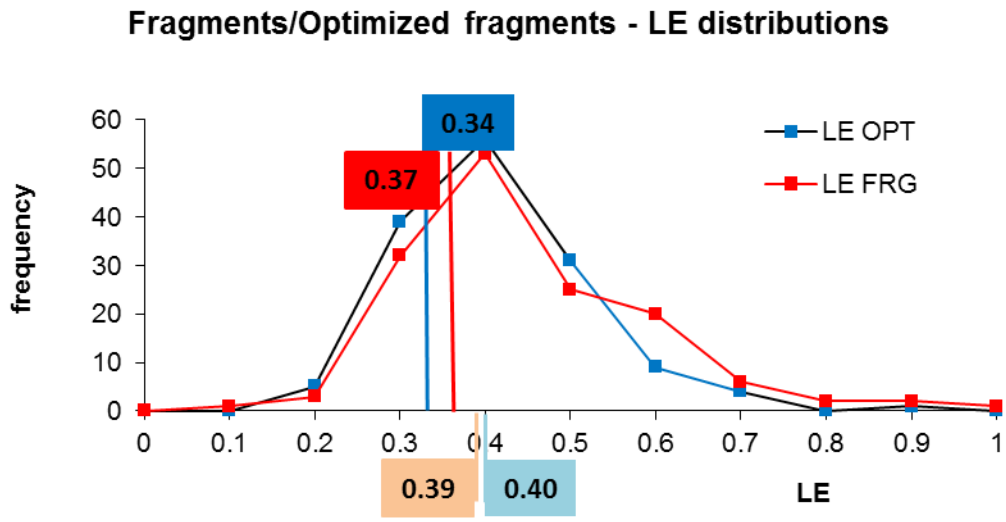

logP change distribution
frequency
25
20
15
10
5
0
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
6.5
7
DlogP

Figure 4

a)


Fragments/Optimized fragments - LE distributions
frequency
LE
LE OPT
LE FRG
0.34
0.37
0.39
0.40
0
10
20
30
40
50
60
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

b)

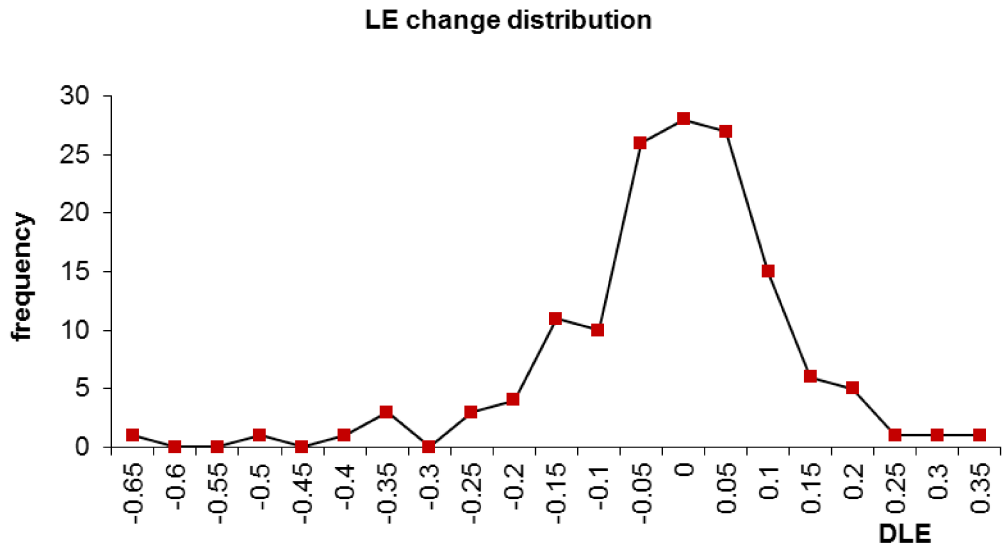
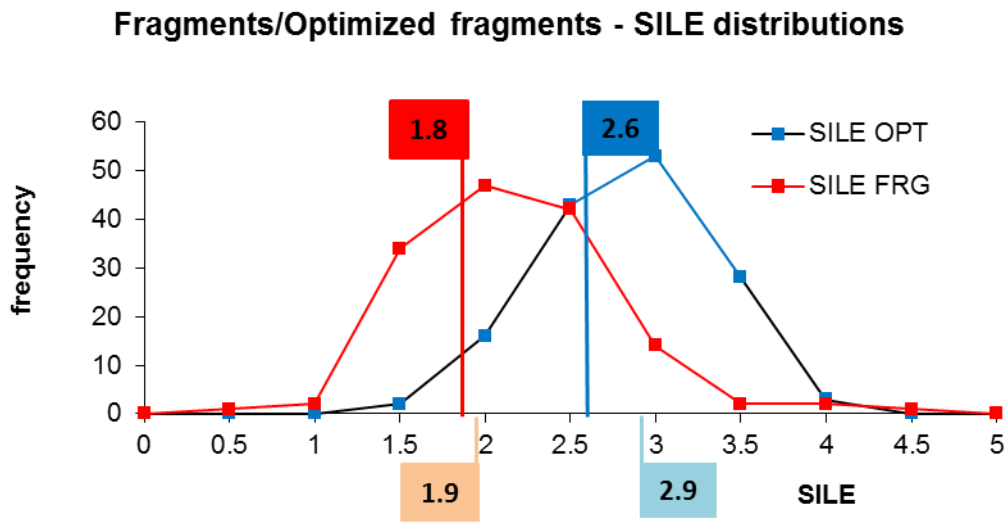

LE change distribution
frequency
DLE
0
5
10
15
20
25
30
-0.65
-0.6
-0.55
-0.5
-0.45
-0.4
-0.35
-0.3
-0.25
-0.2
-0.15
-0.1
-0.05
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35

Figure 5
a)


Fragments/Optimized fragments - SILE distributions
frequency
60
50
40
30
20
10
0
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
1.8
2.6
1.9
2.9
SILE
SILE OPT
SILE FRG

b)

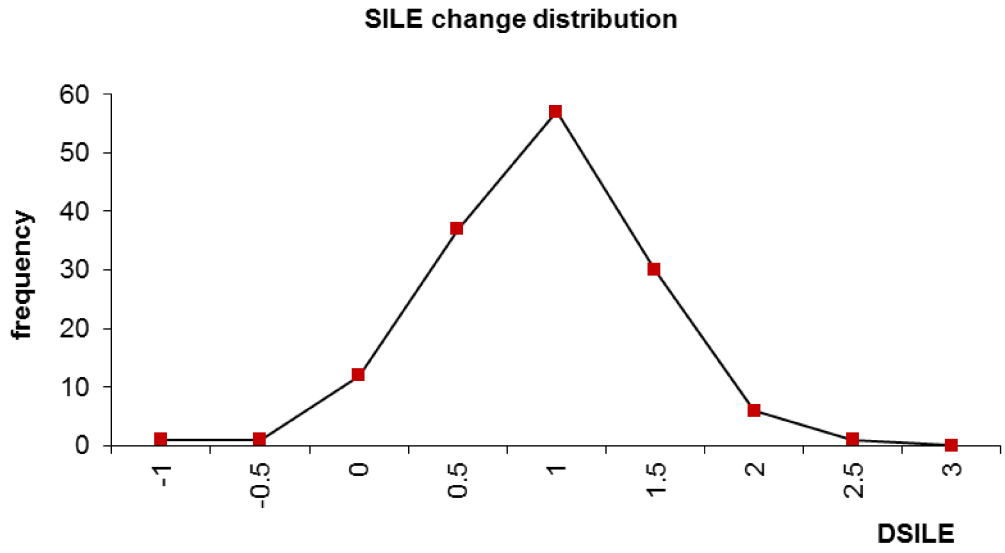
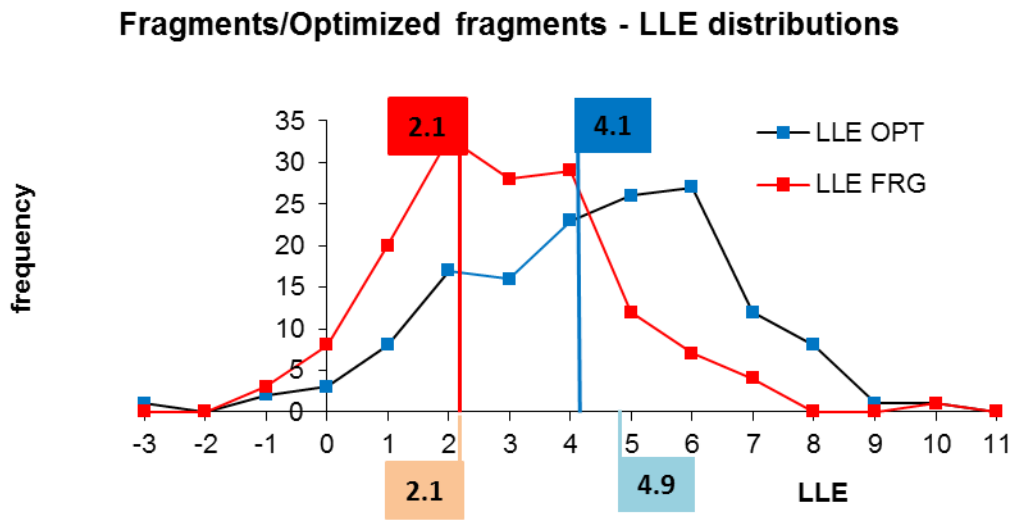

SILE change distribution
frequency
60
50
40
30
20
10
0
-1
-0.5
0
0.5
1
1.5
2
2.5
3
DSILE

Figure 6
a)


Fragments/Optimized fragments - LLE distributions
frequency
35
30
25
20
15
10
5
0
2.1
4.1
LLE OPT
LLE FRG
-3 -2 -1 0 1 2 3 4 5 6 7 8 9 10 11
2.1
4.9
LLE

b)

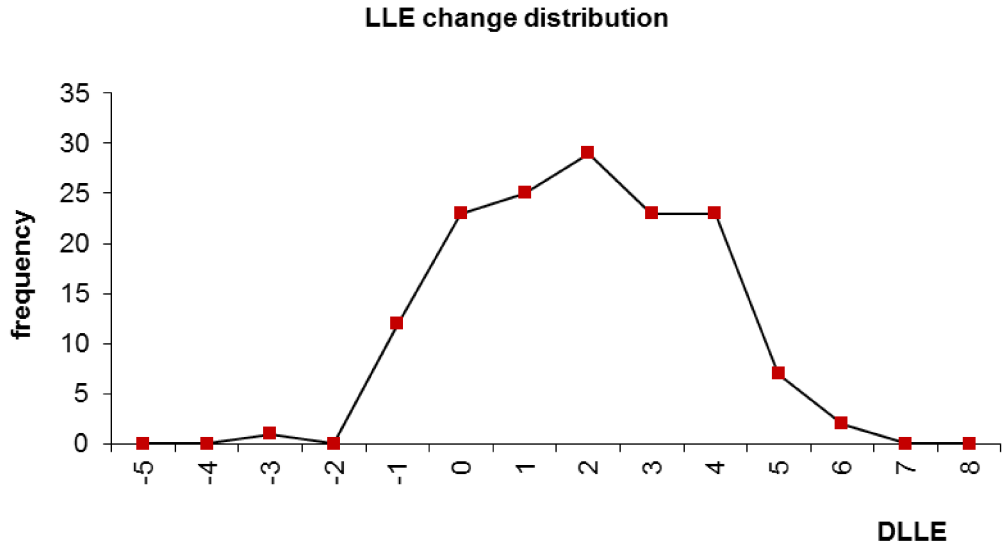
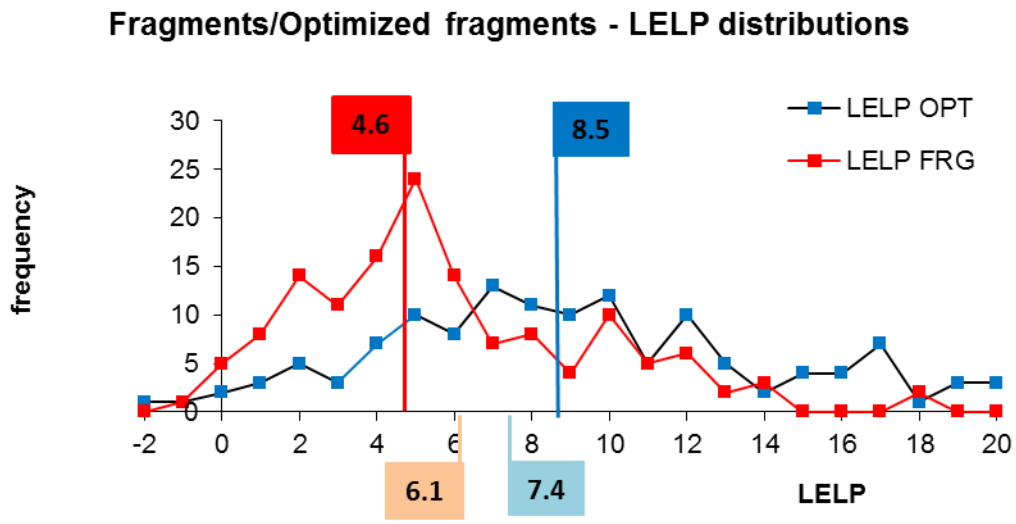

LLE change distribution
frequency
35
30
25
20
15
10
5
0
-5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8
DLLE

Figure 7
a)


Fragments/Optimized fragments - LELP distributions
frequency
LELP OPT
LELP FRG
4.6
8.5
6.1
7.4
LELP
30
25
20
15
10
5
0
-2
0
2
4
6
8
10
12
14
16
18
20

b)

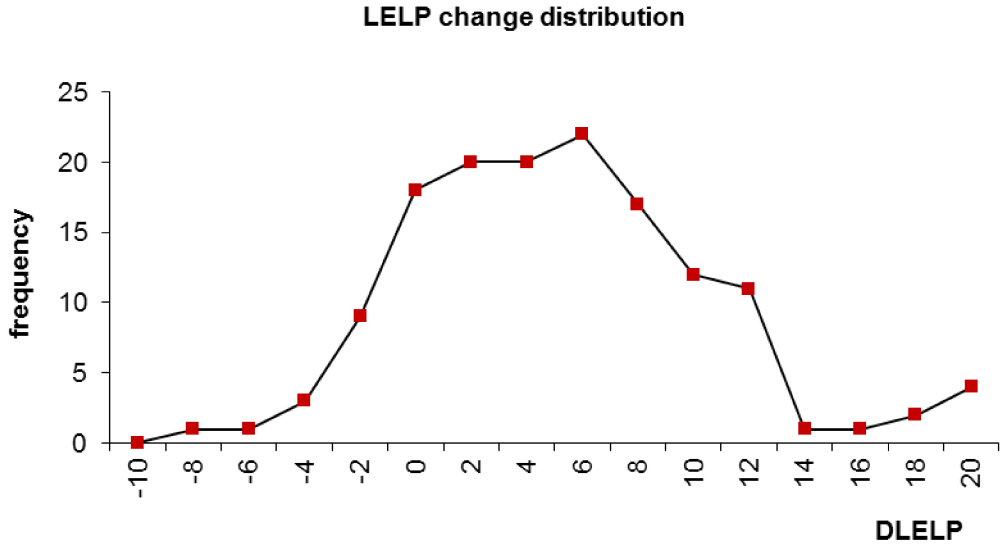

LELP change distribution
frequency
DLELP
25
20
15
10
5
0
-10
-8
-6
-4
-2
0
2
4
6
8
10
12
14
16
18
20

Figure 8

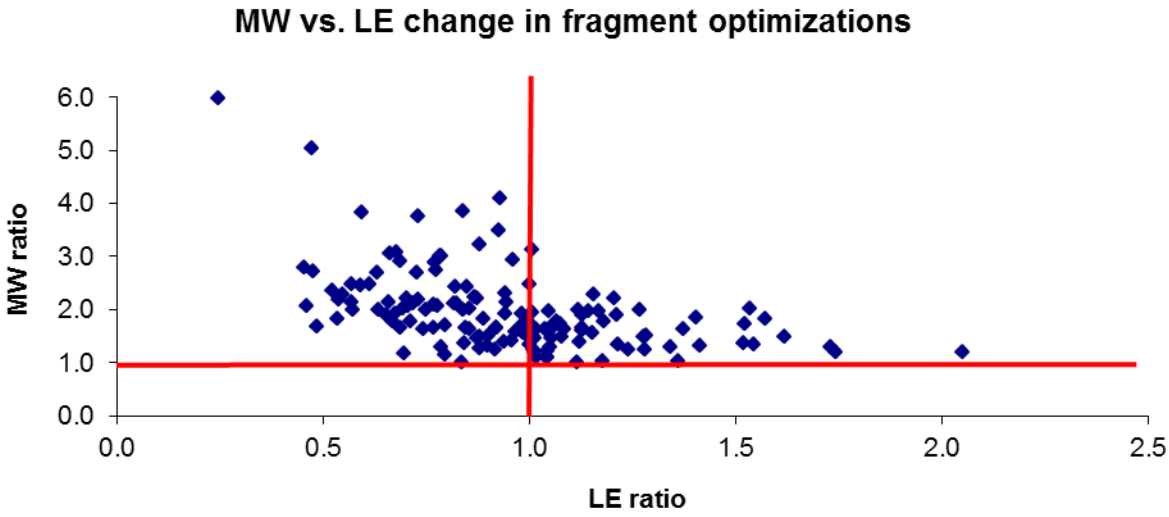

MW vs. LE change in fragment optimizations
MW ratio
6.0
5.0
4.0
3.0
2.0
1.0
0.0
0.0
0.5
1.0
1.5
2.0
2.5
LE ratio

Figure 9

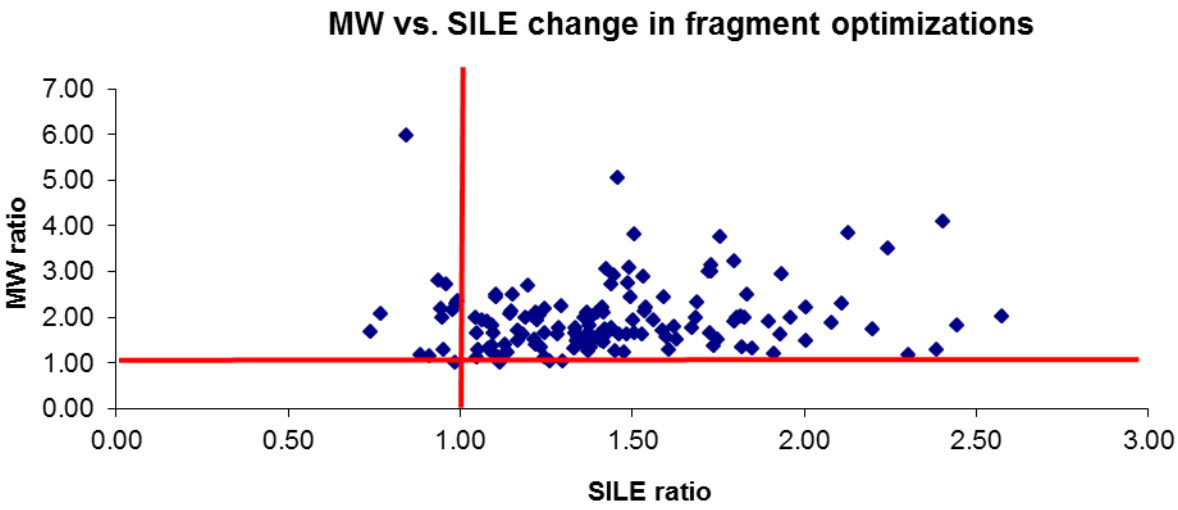

MW vs. SILE change in fragment optimizations
MW ratio
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
0.00
0.50
1.00
1.50
2.00
2.50
3.00
SILE ratio

Figure 10

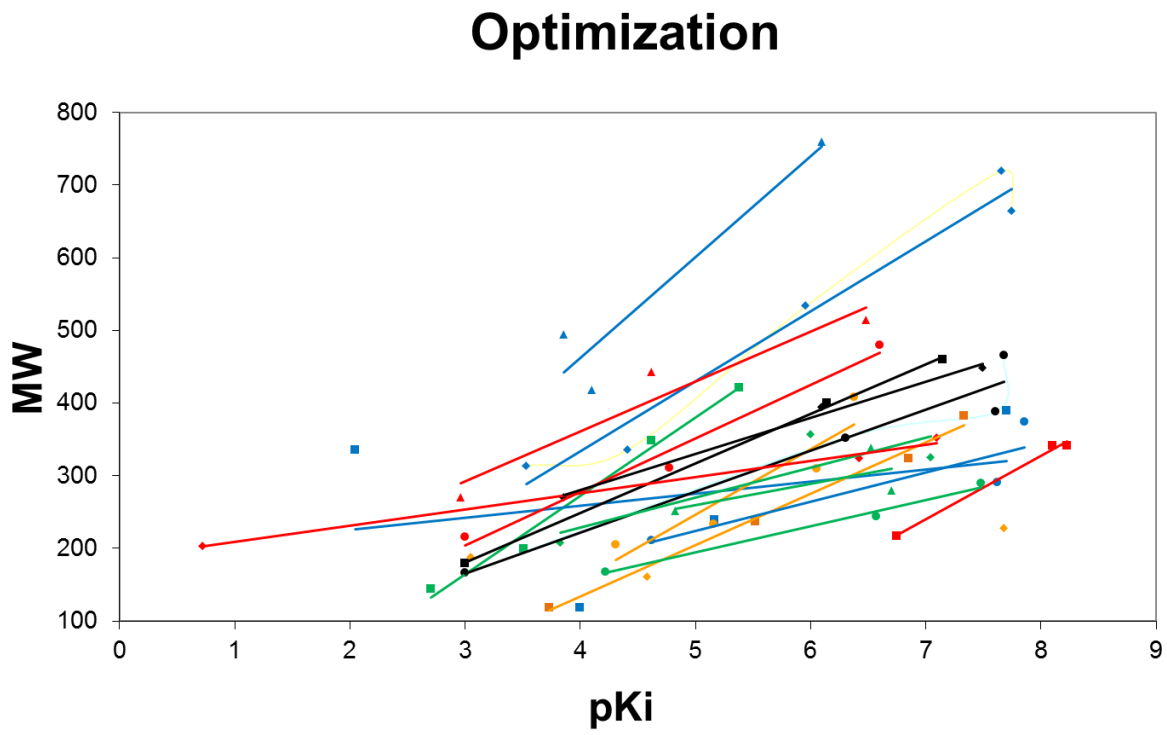

Optimization
MW
pKi
800
700
600
500
400
300
200
100
0
1
2
3
4
5
6
7
8
9

Figure 11

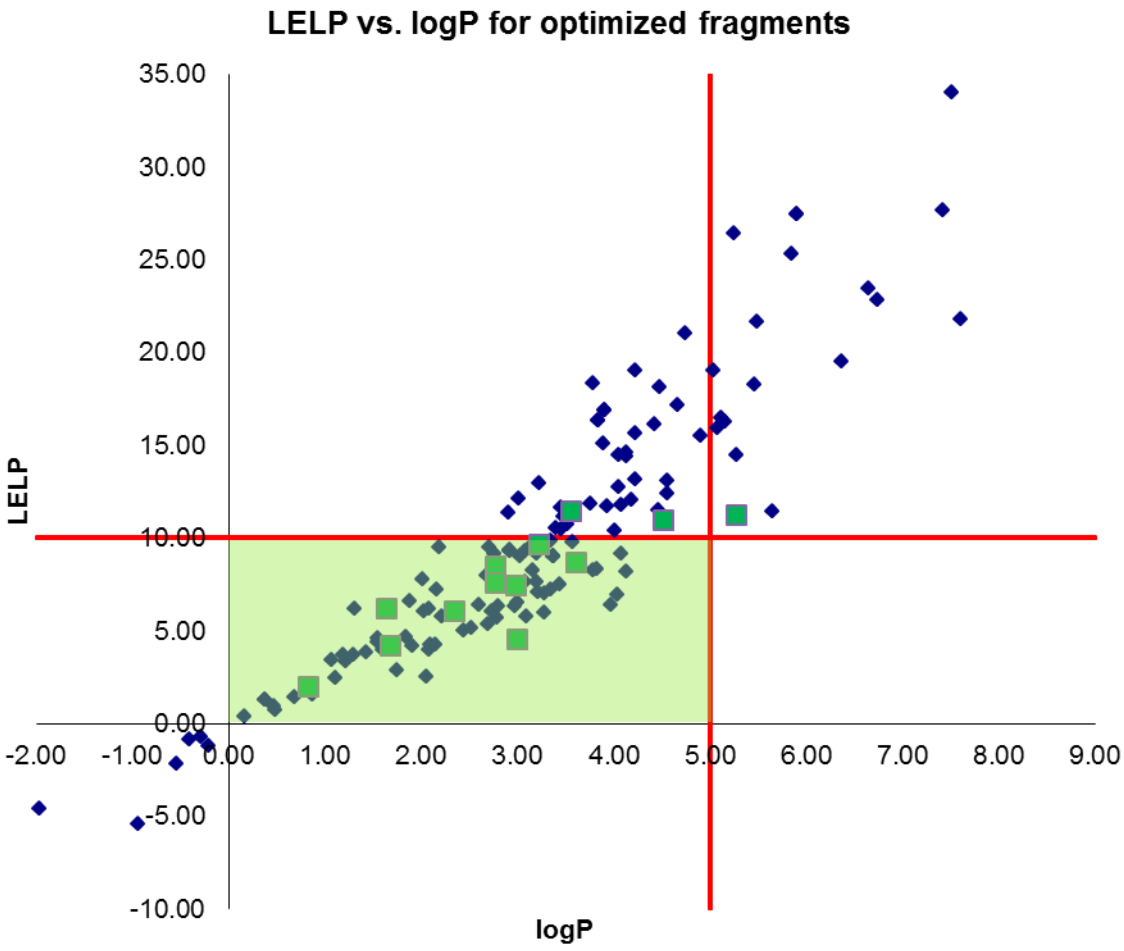

LELP vs. logP for optimized fragments
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
-5.00
-10.00
LELP
-2.00
-1.00
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
logP

Table of Contents graphic.