Binding thermodynamics discriminates fragments from drug-like compounds. A thermodynamic description of fragment based drug discovery

Glyn Williams[1], György G. Ferenczy[2], Johan Ulander[3] and György M. Keserű[2]*

[1]Astex Pharmaceuticals, 436 Cambridge Science Park, Milton Road, Cambridge CB4 0QA, U.K.

[2]Medicinal Chemistry Research Group, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Magyar tudósok krt. 2, H-1117 Budapest, Hungary

[3]CVMD Innovative Medicines, AstraZeneca R&D Mölndal, S-43183 Mölndal, Sweden

Corresponding author:

Prof. György Miklós Keserű

Medicinal Chemistry Research Group,

Research Centre for Natural Sciences,

Hungarian Academy of Sciences

Magyar tudósok körútja 2. H-1117 Budapest, Hungary

Phone: +36-1-3826821

E-mail: gy.keseru@ttk.mta.hu

Web: http://medchem.ttk.mta.hu/index.html

Keywords

fragment based drug discovery, binding thermodynamics, more enthalpy driven binding, druglike molecules, fragments, isothermal titration calorimetry

Teaser

A large thermodynamic dataset from Astex, AstraZeneca, Pfizer and academic labs which includes fragment-protein interactions demonstrates that, when compared to many traditional drug-like compounds, fragments bind more enthalpically to their protein targets.

Abstract

Small is beautiful – reducing the size and complexity of chemical starting points for drug design allows better sampling of chemical space, reveals the most energetically important interactions within protein binding sites and can lead to improvements in the physico-chemical properties of the final drug. The impact of fragment based drug discovery (FBDD) on recent drug discovery projects and our improved knowledge of the structural and thermodynamic details of ligand binding has prompted us to explore the relationships between ligand binding thermodynamics and FBDD. Information on binding thermodynamics can give insights into the contributions to protein-ligand interactions and may therefore be used to prioritise compounds with a high degree of specificity in forming key interactions.

Ligand size and binding thermodynamics

The maximal available binding affinity depends on ligand size and this observation appears to support the medicinal chemistry practice that adds new functional groups in order to improve affinity. In contrast, it has been claimed that the maximal available enthalpy gain decreases with increasing ligand size or, viewed alternatively, that medicinal chemistry optimisation has traditionally tended to enhance affinity mainly for entropic reasons [1-3]. However, this may be an over-simplification when considering the effects of small structural changes between similar ligands [4]. While high favourable enthalpy is accompanied by high affinity in the case of small ligands, this is not necessarily true for large ligands, where higher affinity compounds bind typically with lower enthalpy gain and further improvement in binding enthalpy is often achieved at the expense of affinity. These observations are based on the analysis of large, publically-available isothermal titration calorimetry (ITC) data sets, such as the Scorpio (URL: http://scorpio.biophysics.ismb.lon.ac.uk/scorpio.html), and BindingDB (URL: http://www.bindingdb.org/bind/index.jsp) databases. Since binding enthalpy broadly reflects the overall quality of protein-ligand interactions, the opposite size dependence of affinity and enthalpy may have fundamental consequences for drug-discovery practices, including the identification of chemical starting points and their optimization: (i) more enthalpic binders may prove to be more advantageous starting points for medicinal chemistry optimizations; (ii) enthalpic optimization may be more beneficial during the early phase of optimizations; and (iii) smaller compounds may be more likely to bind enthalpically than larger compounds.

The demonstrated success of fragment based drug discovery programs and the proposed consequences prompted us to investigate the binding of fragment-size compounds by analysing their complexes in the PDB [1]. It was found that fragments often form a small number (two on average) of near-to-optimal geometry H-bonds. This is a consequence of their small size that allows them to form good quality H-bonds with low steric constraints. These usually contribute decisively to the binding free energy and this contribution is predominantly enthalpic, overcoming the loss of ligand rigid-body entropy that is also associated with binding. The complexity model of Hann [5] supports the view that ligands are able to form a limited number of optimal interactions and that their number does not increase with increasing ligand size and complexity. Indeed, it was found that the burial of 50-100 $Å^2$ polar surface area that can be associated with the formation of two H-bonds results in a significant enthalpic benefit but this benefit does not increase with increasing polar surface area burial [6].

Fragments usually bind to a confined region within a larger protein binding site and this is identified as the hot spot. Hot spots are energetically important regions of the binding site; they are able to

bind a diverse set of small organic compounds [7,8], and they bind fragments in a way that the extension of the fragments to larger compounds does not affect the binding pose [9].

A consequence of the small size of fragments is that their binding to the hot spot disturbs the water network to a lesser extent and in a different manner than do large ligands. A particular feature of hot spots is that they are often associated with water molecules which have unfavourable excess entropy [10]. Using inhomogeneous fluid solvation theory Huggins estimated the enthalpic and entropic contributions of individual water molecules in 19 protein cavities of 5 proteins [11] and concluded that the contribution of entropic penalty of water molecules in protein cavities might be small to the free energy. These observations suggest that the small number of such water molecules released by fragment-binding is usually unable to turn the entropy balance. This is well illustrated by the binding of a series of fragments to carbonic anhydrase where the release of water molecules plays a decisive role in determining the relative enthalpy and entropy content of binding. Nevertheless, the overall contribution of water release to the binding enthalpy is a fraction of the observed enthalpy gain and does not affect the substantially favourable enthalpy [12].

Apolar desolvation contributes to the binding of larger ligands and it was shown that the burial of ~20 $Å^2$ apolar Connolly surface upon binding typically leads to ~1 $kJ.mol^{-1}$ free energy gain [6]. While this dominantly entropic contribution is significant for large ligands it is less important for fragments owing to their small size and buried surface. Moreover, the rigid body entropy loss upon ligand-protein binding amounts to approximately 15-20 $kJ.mol^{-1}$ [13] that must be compensated before entropically favoured binding is observed. This latter entropy loss only slightly increases with ligand size and for this reason its contribution is relatively more important for fragments than for larger ligands.

The observations that fragments can achieve higher favourable binding enthalpy than larger ligands; that they can form optimal geometry H-bonds in the protein hot spot without incurring a large apolar desolvation penalty; and that this occurs without significantly disturbing the water network in the binding site, strongly suggest that fragments would be expected to bind enthalpically.

Experimental thermodynamic data of fragment binding

There is a large body of experimental evidence that is in line with the expectations discussed above; fragments bind to proteins with favourable enthalpy. These data come most abundantly from direct measurements of enthalpy. With recent methodological and technical developments [14], Isothermal Titration Calorimetry (ITC) experiments have become feasible for low affinity complexes, allowing the accurate measurement of significant enthalpies ($|\Delta H_{binding}|$>5$kJ.mol^{-1}$) for binding affinities in the range 100 µM < K_d < ~1 mM ('low-c titrations') and 10 nM < K_d < 100 µM for direct titrations [15].

First we investigated the binding thermodynamics profile of fragments reported in the biomedical literature. Binding thermodynamic data of this set of diverse 284 fragments are shown in Figure 1 panels A and B, and indicate that the majority of the fragments bind with a favourable enthalpy change [1]. The few exceptions found in the public dataset are all charged compounds and the observed entropy dominated binding is in line with the large enthalpic penalty of de-solvation for ions (Table 1).

Figure 1 panel A gives a traditional representation of this thermodynamic binding data, where ΔH is plotted against $-T\Delta S$. Large areas of such graphs are empty, corresponding to complexes where the

affinities are too weak to measure ($\Delta G>0$ kJ.mol^{-1}) or too tight to achieve with typical non-covalent interactions ($\Delta G<-60$kJ.mol^{-1}). Data in Figure panel B and those in all subsequent figures are shown as ΔG versus $\Delta H+T\Delta S$. This change of axes leads to a 45° rotation of the data when compared to the more usual representation of ΔH versus $-T\Delta S$ in panel A. The quantity $(-1/T).(\Delta H + T\Delta S)$ has a physical meaning, corresponding to the difference between the entropy created in the system (ΔS) and outside the system ($-\Delta H/T$), for a closed system undergoing a spontaneous change. Compounds with both favourable enthalpy and entropy appear in the lower-middle triangle while those with unfavourable enthalpy are above the right diagonal and those with unfavourable entropy are above the left diagonal. Areas which contain no data points can be excluded by restricting the ΔG axis scale.

This representation also suggests the use of the ratio $(\Delta H+T\Delta S)/\Delta G$ as a measure of the enthalpic driving force. This quantity is zero when enthalpy and entropy contribute equally to the free energy of binding and is positive for enthalpy-driven binding ($\Delta H < -T\Delta S$). Its value exceeds 1 when enthalpy must overcompensate an unfavourable binding entropy (compounds above the left diagonal in Figure panel B). By contrast, a negative value of the ratio indicates an entropy-driven binding while values smaller than -1 correspond to a favourable entropy that overcompensates an unfavourable enthalpy (compounds above the right diagonal of Figure 1 panel A).

Consequently, we define the dimensionless ratio $(\Delta H+T\Delta S)/\Delta G$ as the Enthalpy-Entropy Index ($I_{E\text{-}E}$) and use it here to indicate the enthalpy content of binding. Its advantageous feature is that it is normalised by the free energy ΔG ($=\Delta H-T\Delta S$), and so it can be used to compare compounds with millimolar to nanomolar binding affinities, during the course of a hits-to-leads optimisation.

Thermodynamic binding data for a larger set of carefully selected fragments has been obtained at Astex and Astra-Zeneca, in the course of their drug discovery programs and is shown in Figure 1 panel C. The Astex data shown in Figure 1 panel C includes a total of 782 ITC measurements obtained from fragment screening hits, optimized fragment hits, fragment-derived leads and optimized leads, binding to a diverse set of 24 target proteins. This dataset complements those from public sources and shows that an overwhelming majority of compounds (98.7%) investigated at Astex bind with favourable enthalpy. The best fit line suggests entropies of binding are, on average, 23 kJ.mol^{-1} less favourable than binding enthalpies. Similar trends were found in a smaller set of fragments identified in AstraZeneca drug discovery programs. These data also demonstrate that fragment hits bind dominantly by favourable enthalpy and this feature can be preserved while they are optimized into leads and drug candidates.

Twelve complexes from the Public dataset, ten complexes of the Astex dataset and one complex from the AstraZeneca dataset display positive (unfavourable) binding enthalpies in Figure 1b and Figure 1 panel C. More information on these complexes is provided in Table 1. It should be noted that data for five of the Astex complexes with unfavourable enthalpy have been omitted from Table I since the data were obtained using a displacement ITC experiment format. In these cases, the measured enthalpies also depend on an accurate knowledge of the binding enthalpy of the competitor ligand and so may contain larger errors than the complexes measured using direct ITC methods.

The ligands in the 18 complexes of Table I have diverse chemical structures and physico-chemical properties that are reasonably representative of a set of drug-like hits and leads, with MW_{ave}= 262Da, $clogP_{ave}$=1.4 and a calculated charge between +1 and -2. Similarly, the protein targets are diverse and span a number of different target classes, containing the substrate and cofactor binding-sites of

enzymes, and the protein-protein and protein-small molecule binding domains of recognition modules.

One ligand, present in Astex5 complex, also binds to a second target in the full Astex set with K_d~200μM and ΔH~-20kJ.mol^{-1} and so its unfavourable binding enthalpy to target B cannot be ascribed to the fragment alone. Conversely, all of the Astex targets in Table 1 form many complexes which have favourable enthalpies and so the observation of an unfavourable enthalpy is also not a property of the target alone.

While the complexes of Table 1 are clearly distinct from the overwhelming majority of complexes described in Figure 1, an understanding of their unusual thermodynamic signatures requires additional information which will likely include changes in the protonation and hydration states of both the ligand and target when the complex is formed. It should also be noted that the ligands described in Table 1 are not necessarily poor starting points for drug design. The ligands contained in the complexes Astex 2 and Astex 4 were identified by fragment screening of the ATPase domain of HSP90 and were starting points for the development of Onalespib (AT13387), a molecule which is currently in Phase II clinical trials for cancer [16, 17]. The ligand of Astex 4 complex corresponds to compound 19 of ref. 14 and has I_{E-E} = -1.5. Addition of just two atoms to give compound 30 of ref. 14, improved its binding affinity by almost 10^4-fold ,its binding enthalpy by 39kJ.mol^{-1} and increased its enthalpic efficiency to I_{E-E} = 0.5. Onalespib itself binds to the HSP90 ATPase domain with an affinity of 0.7nM, an enthalpy of -41kJ.mol^{-1} and I_{E-E} = 0.6 [17]. However, it should be stressed that such behaviour is unusual: as will be shown below, fragment hits generally have values of I_{E-E} close to unity ($-T\Delta S$ ~0) and this value decreases significantly during a typical lead optimisation process.

Further thermodynamic data for several protein targets are available from fragment screening campaigns, from fragment optimizations and from systematic analyses of congeneric series described in the literature. Panel C of Figure 1 shows target specific binding thermodynamics data that include 11 fragments tested against trypsin [18], 16 and 7 fragment hits against Mycobacterium tuberculosis pantothenate synthetase [19] and PLP-dependent transaminase (BioA) [20], respectively, 58 fragments screened against p38alpha [21], 11 fragments target Pseudomonas aeruginosa PqsR protein [22], 20 congeneric fragments binding to human carbonic anhydrase II [23], and 4 fragments measured by direct ITC against thrombin [24]. All of these 127 fragments bind with favourable enthalpy. More recently we compiled [25] a dataset of 138 neutral fragments (94) and drug like compounds (44) acting on 17 targets that showed the pronounced tendency of fragments to bind with more favourable enthalpy and less favourable entropy with respect to drug-like compounds binding to the same targets (Figure 1, panel E).

It is important to note that the compounds in ref. 25 were all evaluated by direct ITC measurement (92% of the compounds in the dataset show K_d values lower than 100 μM), and no displacement experiments were included. A statistical analysis of ΔG, ΔH and $-T\Delta S$ values on this dataset showed statistically significant difference for the enthalpic and entropic components of fragments and drug-like compounds (Mann-Whitney U-test, $p<0.005$). These data also imply that, on average, the relative contribution of the enthalpic component to fragment binding, measured using the Enthalpy-Entropy index, I_{E-E}, is greater than that observed for drug-like compounds. There is a statistically significant difference at the p=0.0009 significance level between the value of I_{E-E} for fragments and drug-like compounds with medians 1.10 and 0.79, respectively (Figure 2).

Similarly to publicly available databases, fragment thermodynamic data from both corporate and academic drug discovery laboratories (a total of almost 1000 data points) collected for a wide variety of targets show that fragments bind with favourable enthalpy. Moreover, the binding enthalpy dominates in the large majority of cases. This clearly distinguishes fragments from larger compounds where such preference for enthalpy dominated binding cannot be observed.

Errors in thermodynamic quantities derived from ITC

In typical ITC experiments used to generate the data described above, a sample of the protein ('the titrand') is contained within a small reaction cell which is thermally insulated from the environment, at the centre of a titration calorimeter. For typical calorimeters and binding experiments, the protein concentration would be 5μM to 10 μM and the cell volume is 0.3ml to 1.5ml. Small volumes of a concentrated solution of the ligand ('the titrant') are then added via a syringe, which also serves to stir the solution, thus ensuring rapid mixing. If the ligand binds to the protein with a non-zero enthalpy, heat is either released or absorbed, leading to a small temperature change in the cell. An electrical heater is used to maintain a constant temperature difference between the reaction cell and a reference cell within the calorimeter, measured using a sensitive thermocouple. The change in heater power required to maintain a fixed temperature difference is then integrated over time and the result corresponds to the heat change on ligand binding in the reaction cell.

Usually, several injections are made to reach a 1:1 stoichiometry of protein and ligand and additional injections are then made to ensure that the protein binding site is saturated. Each injection in the first phase releases a small proportion of the binding enthalpy. For a 1ml cell containing 10 μM of protein with a typical ligand-binding enthalpy of -40kJ.mol^{-1}, each injection releases around 40uJ of heat. To put this in perspective, this is the same amount of heat that would fall on an A4 sheet of paper in 1 second, when illuminated by a 40W bulb, placed nearly 5km away.

It is unsurprising that such calorimetric experiments require sensitive, well-maintained , properly calibrated instruments and precisely prepared solutions. Errors in the molar concentrations of titrant or titrand will result in proportionate errors in measured binding enthalpies (ΔH , kJ.***mol^{-1}***) and dissociation constants (K_d , ***mol.***dm^{-3}). However, since free energies are calculated from the logarithm of K_d, the value of ΔG will contain a smaller percentage error. For example, a 25% error in the concentrations would lead to an error of ~5kJ.mol^{-1} in the calculated molar enthalpy when $\Delta H_{binding}$ = -20kJ.mol^{-1}. However, a 25% error in K_d only causes an error of 0.6kJ.mol^{-1} in ΔG, which is equivalent to a 2% error in $\Delta G_{binding}$ when K_d~1 μM, or a 4% error when K_d~1mM. Entropies (-TΔS) are calculated as the difference between ΔG and ΔH and so the numerical value of the entropic error will closely mirror that of ΔH, with an opposite sign. This correlation of the errors in ΔH and TΔS measured by ITC is separate from the more familiar enthalpy-entropy compensation, in which changes in ΔG usually occur with larger and opposing changes in ΔH and TΔS [2].

Other sources of error or variation must also be recognised and reduced. Heat may be generated simply by the dilution of the titrant into the reaction cell. This heat of dilution can be estimated from injections made after the protein is saturated and must be subtracted from all injection when the data are analysed. Finally changes in the pH or buffer concentration during the course of the titration or between experiments can lead to changes in the protonation state of the protein or the ligand or their weak interactions with ions in solution. Both of these events may be associated with their own heat changes.

The practical effects of these errors on measurements of ΔH were investigated at Astex by comparing replicate ITC data, obtained from independent ITC experiments over the course of 8 drug discovery programs. An initial search of the Astex database revealed 80 ITC datasets that were part of replicate measurements involving 30 unique ligands. The smallest number of replicate titrations was 2 while the largest was 7. The maximum variation in $\Delta H_{binding}$ measured from replicate titrations for each ligand is illustrated in Figure 3.

For 70% of the data shown in Figure 3, the variation in ΔH between replicate titrations is better than 5kJ.mol^{-1}. However, target 2 and target 4 (CDK2) show some variations that are greater than 12kJ.mol^{-1}. In all cases the buffer was unchanged between the replicate titrations. Closer inspection of the database showed that the largest variations between replicate measurements of ΔH involved comparisons of different protein constructs (target 2: long +C-terminal tag vs short + N-terminal tag) or different protein complexes (target 4: CDK2 vs CDK2.cyclinA).

After removal of all data involving comparison of different forms of the target (different constructs, complexes or phosphorylation states), 56 ITC datasets remained that formed true replicate titrations for 22 ligands with 7 protein targets. Target 5 has no true replicate data; the remaining targets contain 2 to 7 replicate titrations.

The maximum variation in $\Delta H_{binding}$ measured from true replicate titrations for each ligand is illustrated in Figure 4. This shows that the maximum variation in ΔH observed between true replicate titrations for any of the 22 complexes was 5.2kJ.mol^{-1} and the average of the maximum variations was 2.3kJ.mol^{-1} (S.D. = 1.8kJ.mol^{-1}). Comparison of Figure 4 with Figure 3 indicates that minor modifications to a protein target such as changes in construct length, post-translational modifications remote from the ligand binding site and formation of additional protein-protein interactions can substantially change the binding enthalpy of small ligands, here by up to 10kJ.mol^{-1}.

Although this analysis has focussed on replicate titrations for which the expected difference in ΔH is zero, it also indicates that, within the full Astex ITC dataset, errors in ΔH measurements should be less than 2.3kJ.mol^{-1} on average, with 68% having errors less than 4.1kJ.mol^{-1} and 94% having errors less than 5.9kJ.mol^{-1}. Note that the majority of complexes that are listed in Table 1 have ΔH values greater than +5.9kJ.mol^{-1} and so their unfavourable binding enthalpies are unlikely to be a result of experimental error.

Concluding remarks

Theoretical considerations and experimental data indicate that fragment binding is typically more enthalpically-driven than the binding of fragment-derived leads and ligands derived by other drug-discovery approaches. The average binding enthalpy, measured by calorimetry for a large diverse set of fragments and targets, is more favourable than the average binding entropy by an amount that agrees well with estimates of the amount of rigid-body entropy that must be surrendered when a freely-rotating ligand in solution forms a geometrically-constrained complex with a large molecule.

Such constraint renders fragments promising starting points for drug discovery programs and creates a thermodynamic rationale for fragment based drug discovery. It is important to remember that increasing the number and strength of 'high quality' interactions such as hydrogen bonds will not necessarily result in an overall gain in enthalpy. The measured binding enthalpy is a net value and the dissection of the individual contributions might be ambiguous. Solute effects, structural flexibility, and cooperativity lead to nonlinear changes in enthalpy and make enthalpy contributions of

individual intermolecular contacts experimentally non-observable. Binding enthalpy and entropy therefore should not be used as direct end points but, together with structural studies and free energy calculations, can deepen our understanding of ligand binding [26]. As one designs larger molecules, contributions from protein and solvent reorganization are expected to be larger. Enthalpic gains are often partially offset by entropic losses as the complex becomes more geometrically constrained, and so changes in the enthalpy and entropy of binding tend to be negatively correlated. Starting from fragments, it is often possible to efficiently improve the affinity and binding enthalpy during early optimization using a combination of structural (primarily X-ray) and thermodynamic (ITC) data.

The association of enthalpic interactions with 'high-quality' binding in drug discovery has a number of origins. Firstly, the optimisation of geometrically constrained interactions favours the use of structure-based design which, when combined with computational methods in iterative cycles of synthesis and testing, restricts the number of chemical targets and improves the efficiency of the process. Secondly, the incorporation of more geometrical constraints into the interaction also provides a simple rationale to predict and possibly improve the selectivity of binding of a ligand to a related set of protein targets. Thirdly, the improvement of affinity using entropic gains is most closely associated with an increase in the lipophilicity of the ligand and the subsequent burial of hydrophobic surface area on binding. This has traditionally led to the generation of drug candidates of high molecular weight ($\geq$500Da) with poor solubility and which also bind non-specifically to cell membranes and transport proteins. This often leads to low efficacy and the observation of metabolic liabilities in clinical trials and has been described as 'molecular obesity' [27].

The small size and low lipophilicity of a well-designed fragment library provide a large operational freedom to optimize fragment hits into development candidates with beneficial physicochemical and ADME properties, as demonstrated by recent FBDD success stories [28]. Furthermore, the requirement for libraries to be tested at high concentrations and therefore to have good aqueous solubilities favours fragments that contain several hydrogen-bond donors and acceptors. Since fragment screening also explores chemical space efficiently and usually generates multiple, chemically diverse hits, the tendency for them to be geometrically-constrained by hydrogen-bonds allows the most common, and hence favourable, interactions with the protein to be identified and probed using new molecules that were not part of the initial screening set. This merging or growing of fragment hits is proving to be a fruitful stimulus for the synthesis of novel, small heterocycles and other chemical entities.

The tendency for fragment binding to be associated with a gain in enthalpy should not be translated into dogma. We have shown examples of two fragments which bind to their target protein, HSP90, with unfavourable enthalpies but favourable entropies, although this situation was rapidly reversed in the course of their initial optimisation using structure-based design. This is not a common occurrence amongst the targets studied to date; the favourable entropy of binding of these fragment hits probably has its origin in the release of several protein-bound water molecules from the active site which form part of a network of hydrogen bonds used to recognise the purine base of its cofactor, ATP. However, these fragments led to the development of a drug-candidate, so such hits should not be overlooked.

When modifications to the ligand improve ΔH but with no change in ΔG, there is still the prospect of further increases in affinity, provided the concurrent change in binding entropy can be addressed. In these circumstances, attention might first focus on restricting the conformation of the free ligand or preserving some flexibility in the protein-ligand complex.

In drug design, optimisation will require compromises between, for example, potency, safety, cellular activity and pharmacodynamic efficacy. Compounds that interact with high specificity with their target will nevertheless stand the best chance of success. A key feature of fragment binding is the identification of near-optimal geometries for polar interactions. The better our understanding of the biophysical parameters involved, the better able we will be to guide drug design towards safe and efficacious compounds.

Acknowledgements

The authors are grateful to Chris Murray and Joe Coyle (Astex) for their comments. The data presented in Figure 1 are the work of many diligent scientists from Astex, AstraZeneca and public laboratories and whose contributions are gratefully acknowledged. This work was supported by National Brain Research Program KTIA-NAP-13-1-2013-0001.

References

1. Ferenczy, G. G., Keserű, G. M. (2010) Enthalpic efficiency of ligand binding. *J. Chem. Inf. Mod.* **50**, 1536-1541
2. Ladbury J. E., Klebe G. and Freire E. (2010) Adding calorimetric data to decision making in lead discovery: a hot tip. *Nat. Rev. Drug Discov.* **9**, 23-27
3. Ferenczy, G.G., Keserű, G.M. (2012) Thermodynamics of Fragment Binding. J. Chem. Inf. Mod. 52, 1039-1045
4. Klebe, G. (2015) Applying Thermodynamic Profiling in Lead Finding and Optimisation. Nat. Rev. Drug Disc. 14, 95-110
5. Hann, M.M., Leach, A.R., Harper, G. (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comput. Sci.* **41**, 856-864
6. Olsson, T.S., Williams, M.A., Pitt, W.R., Ladbury, J.E. (2008) The thermodynamics of protein–ligand interaction and solvation: insights for ligand design. *J. Mol. Biol.* **384**, 1002–1017
7. Hajduk, P. J., Huth, J.R., Fesik, S.W. (2005) Druggability indices for protein targets derived from NMR-based screening data. *J. Med. Chem.* **48,** 2518–2525
8. Vajda, S,, Guarnieri, F. (2006) Characterization of protein-ligand interaction sites using experimental and computational methods. *Curr. Opin. Drug Discov. Devel.* **9**, 354–362
9. Kozakov, D., Hall, D. R., Jehle, S., Luo, L., Ochiana, S. O., Jones, E. V., Pollastri, , Allen, K. N., Whitty, A., Vajda, S. (2015) Ligand deconstruction: Why some fragment binding positions are conserved and others are not. *Proc. Natl. Acad. Sci. USA,* **112,** E2585
10. Ichihara, O., Shimada, Y., Yoshidome, D. (2014) The importance of hydration thermodynamics in fragment-to-lead optimization. *ChemMedChem* **9,** 2708-17
11. Huggins, D. J. Quantifying the Entropy of Binding for Water Molecules in Protein Cavities by Computing Correlations. (2015) *Biophysical J.* **108**, 928–936
12. Breiten, B., Lockett, M.R., Sherman, W., Fujita, S., Al-Sayah, M., Lange, H., Bowers, C.M., Heroux, A., Krilov, G., Whitesides, G.M. (2013) Water networks contribute to enthalpy/entropy compensation in protein-ligand binding. *J. Am. Chem. Soc.* **35**, 15579-84
13. Murray, W.C., Verdonk, M.L. (2002) The consequences of translational and rotational entropy lost by small molecules on binding to proteins. *J. Comput.-Aided Mol. Design* **16,** 741-753

14. Torres, F.E., Recht, M.I., Coyle, J.E., Bruce, R.H., Williams, G. (2010) Higher throughput calorimetry: opportunities, approaches and challenges. *Curr. Opin. Struct. Biol.* **20**, 598–605
15. Krimmer, S.G., Klebe, G. (2015) Thermodynamics of protein–ligand interactions as a reference for computational analysis: how to assess accuracy, reliability and relevance of experimental data. *J. Comp. Aided Mol. Des.* **29**, 867-883
16. Murray, C.W., Carr, M.G., Callaghan, O., Chessari, G., Congreve, M., Cowan, S., Coyle, J.E., Downham, R., Figueroa, E., Frederickson, M., Graham, B., McMenamin, R., O'Brien, M. A., Patel, S., Phillips, T., Williams, G., Woodhead, A.J., Woolford, A.J-A. (2010) Fragment-based Drug Discovery applied to HSP90. Discovery of Two Lead Series with High Ligand Efficiency. *J. Med. Chem.* **53**, 5942-5955
17. Woodhead, A.J., Angove, H., Carr, M.G., Chessari, G., Congreve, M., Coyle, J.E., Cosme, J., Graham, B., Day, P.J., Downham, R., Fazal, L., Feltell, R., Figueroa, E., Frederickson, M., Lweis, J., McMenamin, R., Murray, C.W., O'Brien, M.A., Parra, L., Patel, S., Phillips, T., Rees, D.C., Rich, S., Smith, D-M., Trewartha, G., Vinkovic, M., Williams, B., Woolford, A.J-A. (2010) Discovery of (2,4-Dihydroxy-5-isopropylphenyl)-[5-(4-methylpiperazin-1-ylmethyl)-1,3-dihydroisoindol-2-yl]methanone (AT13387), a Novel Inhibitor of the Molecular Chaperone Hsp90 by Fragment Based Drug Design. *J. Med. Chem.* **53**, 5956-5969
18. Dolezal, O., Doughty, L., Hattarki, M.K., Fazio, V.J., Caradoc-Davies, T.T., Newman, J., Peat, T.S. (2013) Fragment Screening for the Modelling Community: SPR, ITC, and Crystallography. *Austral. J. Chem.* **66**, 1507-1517
19. Silvestre, H.L., Blundell, T.L., Abell, C., Ciulli, A. (2013) Integrated biophysical approach to fragment screening and validation for fragment-based lead discovery. *Proc. Natl. Acad. Sci. U S A.* **110**, 12984-9
20. Dai, R., Geders, T.W., Liu, F., Park, S.W., Schnappinger, D., Aldrich, C.C., Finzel, B.C. (2015) Fragment-based exploration of binding site flexibility in Mycobacterium tuberculosis BioA. J. Med. Chem. **58**, 5208-17
21. Neumann, L., von König, K., Ullmann, D. (2011) HTS reporter displacement assay for fragment screening and fragment evolution toward leads with optimized binding kinetics, binding selectivity, and thermodynamic signature. *Methods Enzymol.* **493**, 299–320
22. Zender, M., Klein, T., Henn, C., Kirsch, B., Maurer, C.K., Kail, D., Ritter, C., Dolezal, O., Steinbach, A., Hartmann, R.W. (2013) Discovery and biophysical characterization of 2-amino-oxadiazoles as novel antagonists of PqsR, an important regulator of Pseudomonas aeruginosa virulence. J. Med. Chem. **56**, 6761-74
23. Scott, A.D., Phillips, C., Alex, A., Flocco, M., Bent, A., Randall, A., O'Brien, R., Damian, L., Jones, L.H. (2009) Thermodynamic Optimisation in Drug Discovery: A Case Study using Carbonic Anhydrase Inhibitors *ChemMedChem* **4**, 1985–1989
24. Rühmann, E., Betz, M., Fricke, M., Heine, A., Schäfer, M., Klebe G. (2015) Thermodynamic signatures of fragment binding: Validation of direct versus displacement ITC titrations. *Biochim. Biophys. Acta* **1850**, 647-56
25. Ferenczy, G.G., Keserű, G.M. (2016) On the enthalpic preference of fragment binding. *Med. Chem. Commun.*, **7**, 332-337
26. Geschwindner, S., Ulander, J., Johansson, P. (2015) Ligand Binding Thermodynamics in Drug Discovery: Still a Hot Tip? *J. Med. Chem.* **58**, 6321-35
27. Hann, M.M. (2011) Molecular Obesity, Potency and Other Addictions in Drug Discovery. *Med. Chem. Commun.* **2**, 349-355

28. Erlanson, D.A., Fesik, S.W., Hubbard, R.E., Jahnke, W., Jhoti, H. (2016) Twenty years on: the impact of fragments on drug discovery *Nat. Rev. Drug Disc.* **15**, 605-609

Figure legends

Figure 1 Binding thermodynamics data for fragments

Panel A: $\Delta H_{binding}$ vs $-T\Delta S_{binding}$ for a structurally diverse set of fragments available from the Scorpio, PDBCal and BindingDB databases

Panel B: ($\Delta H_{binding}$+ $T\Delta S_{binding}$) vs $\Delta G_{binding}$ for the same set of structurally diverse fragments available from Scorpio, PDBCal and BindingDB databases. Note that the change of axes corresponds to a 45 degree rotation of the data in panel A and would allow areas containing no data points to be discarded (ΔG<-60kJ.mol-1 and ΔG>0kJ.mol-1).

panel C: Fragment thermodynamics data from drug discovery programs (blue: Astex, red AstraZeneca)

panel D: Fragments from screening efforts against different targets (green: pantothenate synthetase, light blue: P38alpha, orange: carbonic anhydrase, red: trypsin, black: PLP-dependent transaminase, light green: thrombin, dark red: Pqsr)

panel E: Enthalpic and entropic components of binding for complexes of neutral fragments and drug-like molecules

Figure 2 Distribution and statistics of binding enthalpy (ΔH), binding entropy ($-T\Delta S$) and ($\Delta H+T\Delta S$)/ΔG for neutral fragments and drug-like compounds measured by direct ITC experiments. The analysis considered 94 fragments and 44 drug-like compounds acting on 17 protein targets [25]. Mann-Whitney U-test was applied to test the difference between fragments and drug-like compounds. The results show that fragments bind with more favourable enthalpy (p=0.0001) and less favourable entropy (p=0.0016) with respect to drug-like compounds. Furthermore, the scaled difference between enthalpy and entropy of binding (($\Delta H+T\Delta S$)/ΔG) demonstrates the increased importance of enthalpy gain for fragment binding. The box plots show the median within the box of the 1st and 3rd quartiles together with the range of non-outlier data defined as 1.5 fold the interquartile range around the median.

Figure 3 Variation in $\Delta H_{binding}$ for apparent replicate titrations within the Astex database. 80 ITC datasets contribute to this comparison of complexes between 8 protein targets and 30 ligands. The average variation between these apparent replicates is 4.7kJ.mol^{-1}

Figure 4 Variation in $\Delta H_{binding}$ for true replicate titrations within the Astex database. The average variation between these true replicates is 2.3kJ.mol^{-1}. Target numbering is the same as in Figure 3. Target 5 had no true replicate data and is not present in this figure.

Table 1 Properties of ligands and targets in Figure 1 (panels B and C) which have unfavourable binding enthalpies

Authors

Glyn Williams

Glyn Williams joined Astex in 2001, to develop and apply biophysical methods to fragment-based screening and drug design. This work has involved the use of Nuclear Magnetic Resonance, Isothermal Titration Calorimetry and native Mass Spectrometry and also contributed to the development the Astex fragment library. He is the Vice President, Biophysics of Astex Pharmaceuticals. Previously Glyn had spent 11 years with Roche UK where he was responsible for biological NMR and analytical Mass Spectrometry. After obtaining his degree and doctorate (DPhil.) in Chemistry from the University of Oxford, Glyn held fellowships and lectureships in Inorganic and Bio-Inorganic chemistry at the Universities of Oxford, Sydney and London from 1983 to 1990.

György Ferenczy

György G. Ferenczy received his PhD in computational chemistry from the Eötvös University of Budapest. Following postdoctoral research at the University of Oxford, UK and at the University of Nancy, France, he worked as a computational chemist and as a group leader first at Gedeon Richter (Budapest) and later at Sanofi (Budapest and Strasbourg). Since 2012, he is a senior research fellow at the Semmelweis University and from 2013 at the RCNS of the Hungarian Academy of Sciences. His research interest includes the development and application of computational tools for extended biochemical systems and studying molecular interactions relevant to drug discovery.

Johan Ulander

Johan Ulander currently works as Associate Principal Scientist in the computational chemistry section at Cardiovascular and Metabolic Diseases (CVMD) at AstraZeneca R&D Göteborg, Sweden. Prior to joining AstraZeneca he did post-doctoral research at University of California, San Diego (USCD) and University of Houston. He received his PhD in theoretical physical chemistry from Gothenburg

University and has a B.Sc. in Molecular Biology from the University of Umeå, Sweden. He has over 10 years of experience in drug discovery from early stage hit and target identification to late stage drug optimization. His interests include theoretical biophysics and statistical mechanics with applications in drug design, pharmacokinetics and pharmacodynamics.

György M. Keserű

György M. Keserű obtained his Ph.D. at Budapest, Hungary and joined Sanofi-Aventis CHINOIN, heading a chemistry research lab. He moved to Gedeon Richter in 1999 as the Head of Computer-Aided Drug Discovery. Since 2007, he was appointed as the Head of Discovery Chemistry at Gedeon Richter and contributed to the discovery of the antipsychotic Vraylar™ (cariprazine) that has been approved and marketed in the United States from 2016. From 2013, he served as a director general of the Research Center for Natural Sciences (RCNS) at the Hungarian Academy of Sciences. Now he is heading the Medicinal Chemistry Research Group at RCNS. His research interests include medicinal chemistry and drug design. He has published over 180 papers and more than 15 books and book chapters.

Highlights

- A large thermodynamic dataset has been analyzed, containing binding data for 1364 complexes from Astex, AstraZeneca, Pfizer and academic labs.
- We demonstrate that unlike many drug-like compounds in the existing public databases, fragments bind enthalpically to their protein targets.
- This observation provides an insight into the thermodynamic consequences of fragment based drug discovery and suggests a thermodynamic rationale for prioritising compounds which may form key interactions with a high degree of specificity."