Journal of Medicinal Chemistry

Article

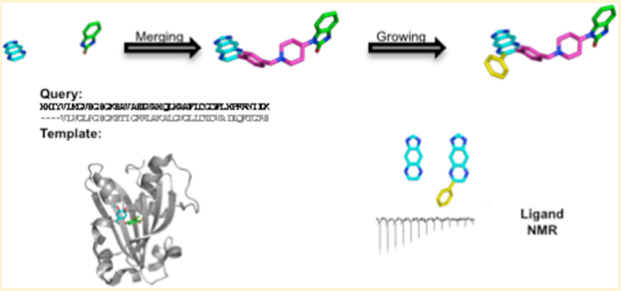
pubs.acs.org/jmc

Toward Rational Fragment-Based Lead Design without 3D Structures

Morkos A. Henen, Nicolas Coudevylle, Leonhard Geist, and Robert Konrat*

Department of Structural and Computational Biology, Max F. Perutz Laboratories, University of Vienna, Vienna Biocenter Campus 5, A-1030 Vienna, Austria

ABSTRACT: Fragment-based lead discovery (FBLD) has become a prime component of the armamentarium of modern drug design programs. FBLD identifies low molecular weight ligands that weakly bind to important biological targets. Three-dimensional structural information about the binding mode is provided by X-ray crystallography or NMR spectroscopy and is subsequently used to improve the lead compounds. Despite tremendous success rates, FBLD relies on the availability of high-resolution structural information, still a bottleneck in drug discovery programs. To overcome these limitations, we recently demonstrated that the meta-structure approach provides an alternative route to rational lead identification in cases where no 3D structure information about the biological target is available. Combined with information-rich NMR data, this strategy provides valuable information for lead development programs. We demonstrate with several examples the feasibility of the combined NMR and meta-structure approach to devise a rational strategy for fragment evolution without resorting to highly resolved protein complex structures.



INTRODUCTION

Fragment-based lead (drug) discovery (FBLD, FBDD) has emerged as a powerful strategy for drug discovery, and numerous successful programs were reported in which series of compounds have entered clinical trials.[1] Of particular relevance is the fact that FBLD strategies have been shown to provide valid starting points for drug discovery even in cases where conventional high-throughput screens (HTS) have failed. The crucial starting point of FBLD is the identification of small molecule weak binders in the 100−300 Da range. Several biophysical techniques exist, among them NMR spectroscopy has proven itself efficient, to provide reliable quantitative binding information. The identified fragments are subsequently evolved in an iterative manner into larger compounds with higher binding affinities and better target selectivity. Fragment optimization is achieved either by linking fragments (fragment merging) or alternatively by the introduction of additional functional groups using synthetic chemistry approaches (fragment extension or growing). The required chemical information is almost exclusively provided by structural studies using (mostly) X-ray crystallography and/or NMR spectroscopy.[2]

Figure 1 illustrates the individual steps of an FBLD program including definition of a suitable fragment library, biophysical detection of weak binders, and identification of binding mode and fragment evolution. The importance of library quality and the necessity of powerful biophysical techniques to detect weak binders for successful drug discovery programs have been described in many articles.[3] Numerous (successful) examples have been described recently in excellent reviews.[3,4] Common belief is that highly resolved structural information is indispensable for subsequent rational medicinal chemistry optimization. The rationale behind this approach is the notion that the 3D structure of the protein carries an imprint of the molecular nature of its partner molecules. Hence, deciphering the molecular interaction code, i.e., identifying the relationship between molecular parameters of the binding cleft and significant chemical descriptors of the ligand, provides the required chemical details to identify suitable chemical derivation and substitution patterns. While this structure-based strategy already delivered series of active compounds (drug candidates) in clinical trials, the lack of structural information available for protein targets of medical interest limits the general applicability of this powerful approach.

Here we want to address the problem of fragment evolution and discuss strategies for binding mode determination, circumventing the bottleneck of highly resolved protein crystal and/or NMR solution structures. The imperative requirement of high-resolution structural information as a starting point for rational drug development programs was recently put into question.[5] It was demonstrated that it is possible to identify valid starting points for ligand development using a strategy based on our protein meta-structure concept, a novel conceptual framework to analyze protein sequences and identify structural and functional features directly from the primary sequence.[5] The meta-structure parameters reveal protein similarities that are hidden on the primary sequence level and can therefore not be identified by conventional (BLAST)[6] sequence analysis tools. In contrast to conventional structure-based drug design programs, only primary sequence information is required for the protein meta-structure similarity

Received: July 13, 2012
Published: August 13, 2012

   dx.doi.org/10.1021/jm301016m | J. Med. Chem. 2012, 55, 7909−7919

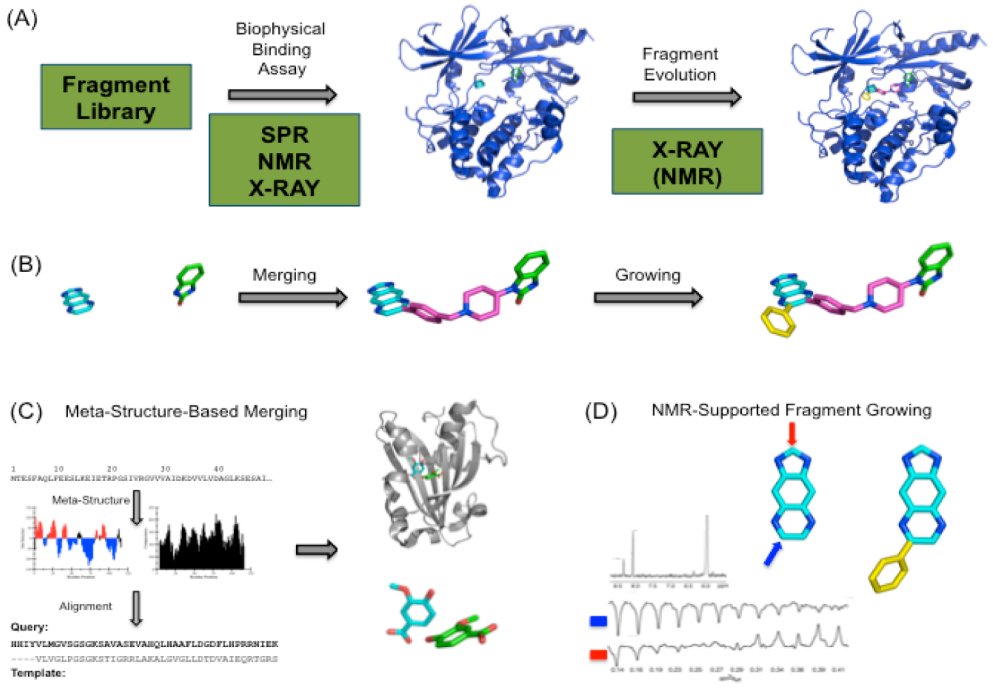


Figure 1. The individual stages of fragment-based lead (drug) design (FBLD). Starting from a suitable chosen small molecule fragment library, biophysical techniques (SPR, NMR, or X-ray) are used to identify weak binders. (A) Structure-based FBLD exploits 3D structural information about ligand binding modes to rationally evolve starting fragments in iterative rounds of optimizations. (B) Fragment evolution is performed by either merging individual fragments binding to different interaction sites or by ligand extension using medicinal chemistry substitution. (C, D) Meta-structure-based fragment-based lead (drug) design strategies for ligand merging (C) and extension (D). (C) Meta-structure homologies are used to discern putative binding modes based on available 3D structure information of the homologue. (D) Suitable sites for ligand derivatization are identified using ligand-based NMR spectroscopy (AFP-NOESY). In this experiment, intraligand $^{1}H-^{1}H$ cross relaxation is monitored as a function of spin lock power. Protons exposed to the solvent exhibit a sign inversion with increasing spin lock power (red). In contrast, protons embedded in hydrophobic clusters (i.e., being part of a dense proton network) display a markedly different behavior (blue) due to spin diffusion. This differential behavior can be used to identify suitable sites for ligand derivatization.

clustering (PMSSC).[5] A first successful application of the PMSSC approach to a protein target from *Thermotoga maritima* demonstrated the applicability to challenging targets.[5] Despite insignificant primary sequence similarities, the PMSSC approach correctly identified homologues of similar enzymatic activity and the chemical structure of possible ligand scaffolds.

NMR spectroscopy has become a powerful experimental tool in drug discovery programs. Although originally introduced as an analytical technique for structure determination of small compounds, its capabilities continued to evolve, and it has now become an indispensable tool in the armamentarium of drug discovery programs. Experiments are available to screen for binders and to analyze protein−ligand interactions (transferred NOE, STD, pumped NOE, waterLOGSY)[7] or, alternatively, to monitor chemical shift or intensity changes (most importantly ^{19}F based detection scheme, such as FAXS[8]). Particularly, in early stages of the FBLD process where often medium-to-weak binders are encountered, NMR spectroscopy is particularly powerful as it is not only a very sensitive detection technique but also provides additional information about binding modes and orientations of bound ligands (mapping of the binding site). Several experiments have thus been developed in the recent past (INPHARMA[9] and SALMON[10]). Most recently, we have devised an NMR pulse sequence for the investigation of protein ligand interactions. In this approach, an adiabatic fast passage pulse is used to probe $^{1}H-^{1}H$ NOEs (AFP-NOESY). We have demonstrated that the presence of spin diffusion leads to significant changes of cross-relaxation pathways and can be used to probe protein−ligand binding interfaces. Given that this methodology is highly sensitive and provides valuable information about potential sites for ligand extensions and/or decoration, widespread applications in fragment-based drug design programs can be anticipated.

Here we present an integrated approach for drug development programs combining fragment-based NMR spectroscopy (AFP-NOESY), meta-structure analysis of protein primary sequences, and general biophysical techniques. The feasibility of the approach is validated with applications to the lipocalin Q83 and the armadillo-repeat region of human β-catenin. This report describes in detail the experimental strategies to ligand merging and ligand extension in FBLD programs and is organized as follows: In the first part, the application to the lipocalin Q83 exemplifies how meta-structure-derived ligand information, NMR spectroscopy, and biophysical techniques can be combined to devise fragment merging strategies without resorting to highly resolved protein complex structures. Identified meta-structure homologies provide valuable ligand information and also offer hints about possible binding modes that subsequently can be used to evolve molecular fragments. Specifically, the binding mode information extracted from the meta-structure could be successfully used to devise a rationale for merging the identified fragments and to obtain an improved ligand with higher binding affinity. The application to the armadillo-repeat region of human β-catenin, a prominent anticancer target, illustrates the potential of ligand-based NMR spectroscopy to provide detailed information about protein ligand binding modes that can be used for rational fragment extension strategies. We demonstrate that measuring intra-

dx.doi.org/10.1021/jm301016m | *J. Med. Chem.* 2012, 55, 7909−7919

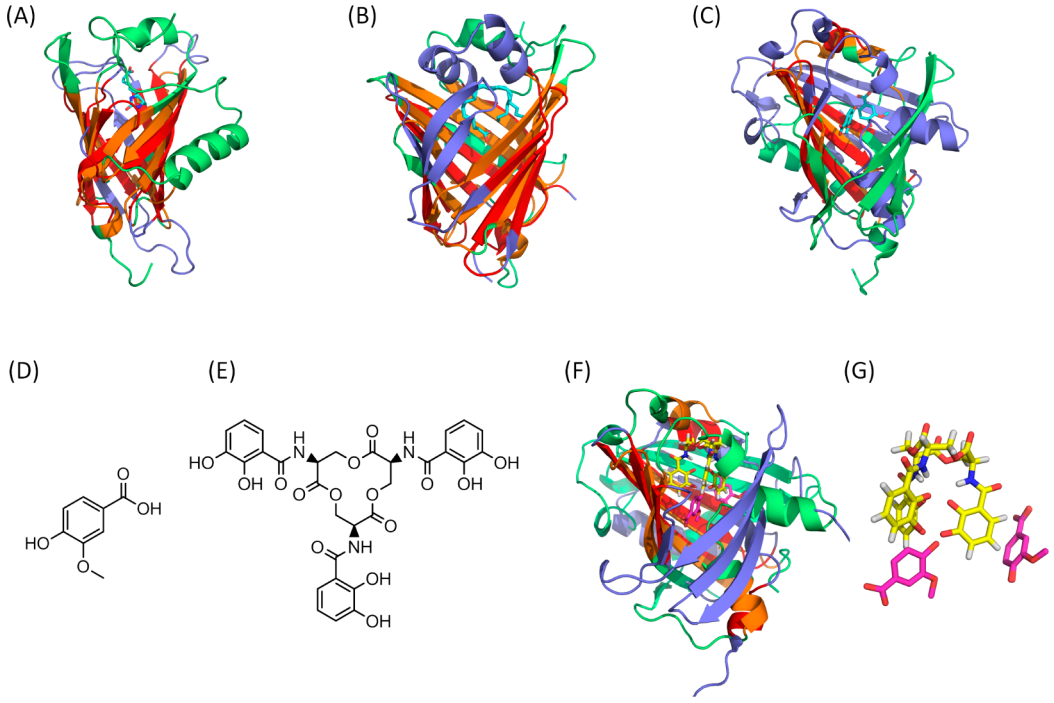


Figure 2. Selection of meta-structure-derived DRUGBANK homologues for the lipocalin protein Q83. (A) Streptavidin, (B) fatty-acid binding protein (FABP), and (C) chorismate lyase. The lipocalin Q83 is shown in green, the DRUGBANK homologues in blue. Regions of structural similarity are indicated in red (Q83) and orange (homologues). Chemical formulas of small molecule ligands for chorismate lyase (D, vanillic acid, VA), and Q83 (E, bacterial siderophore enterobactin). The similar location of enterobactin and vanillic acid in the respective bound states are also indicated in F and G. Figures were prepared using the programs TopMatch (Sippl)[31] and PyMOL (graphics).[32]

ligand cross-relaxation rates during adiabatic fast passage provides unique information about the pharmacophore at the protein ligand interaction site that can be used to identify the binding epitope of the ligand. Knowledge of the binding epitope can be subsequently exploited to identify sites suitable for chemical substitution (ligand extension). Additionally, the application to β-catenin also demonstrates that the meta-structure approach provides valuable additional information for the construction of suitable small molecule (fragment) libraries for ligand screening.

RESULTS

For the meta-structure-based identification of lead compounds we adapted a structure-based strategy developed by the group of Waldmann.[11] Their protein structure similarity clustering (PSSC) approach exploits structural similarities between the therapeutic target of interest and template structures with experimentally verified small molecule ligands. Recently we have shown that the protein meta-structure similarity clustering (PMSSC) approach is a comparably efficient strategy to provide valid starting points even in the absence of structural information about the target protein. The strategy has been described in detail elsewhere.[5] Overall, the strategy is as follows: In a first step, the meta-structure parameters (compactness and secondary structure) are calculated for the therapeutic protein of interest. Second, these data are used to screen (based on pairwise meta-structure alignment) the target sequences of the DRUGBANK database, a public repository of biologically relevant protein targets with experimentally verified inhibitory ligands.[12] It was shown that also in cases where conventional bioinformatical sequence analysis did not reveal any information, the meta-structure approach correctly identified structural homologies that can be exploited for drug design purposes.[5]

Lipocalin Q83. Lipocalin Q83 has first been found to be overexpressed in quail embryonic fibroblasts transformed by the *v-myc* oncogene. The fold of Q83 consists of an eight-stranded antiparallel β-barrel forming a hydrophobic cavity called a "calyx". Q83 is a siderocalin (Scn) as it can capture iron-chelating siderophores with high affinities.[13] Siderocalins participate in the innate immune response by interfering with bacterial siderophore-mediated iron uptake but have also been shown to be involved in many other physiological processes such as inflammation, iron delivery, tissue differentiation, and cancer progression.[14] On the basis of the primary sequence, the meta-structure parameters were calculated and screened against template sequences of the DRUGBANK database as described.[5] The identified hits were scored based on meta-structure similarities. To illustrate the relationship between 3D structural and meta-structural similarities, Figure 2 shows structural superpositions of the identified best scoring homologues with the solution structure of Q83. It can be seen that the identified meta-structure homologues differ in their degrees of structural similarities with Q83, which relates to the fact that meta-structure is not identical to the 3D structure of a protein but goes beyond this conceptual level. The meta-structure homologues streptavidin (Figure 2A) and fatty-acid binding protein (FABP, Figure 2B) clearly show substantial 3D structural similarities. Additionally, the ligand binding sites of Q83, streptavidin, and FABP overlap, which supports the basic assumption that protein meta-structural similarities suggest similar binding modes. On the other hand, homologues such as the *E. coli* chorismate lyase (Figure 2C) only share substructural motifs (e.g., one-half of the β-barrel with Q83).

dx.doi.org/10.1021/jm301016m | J. Med. Chem. 2012, 55, 7909−7919

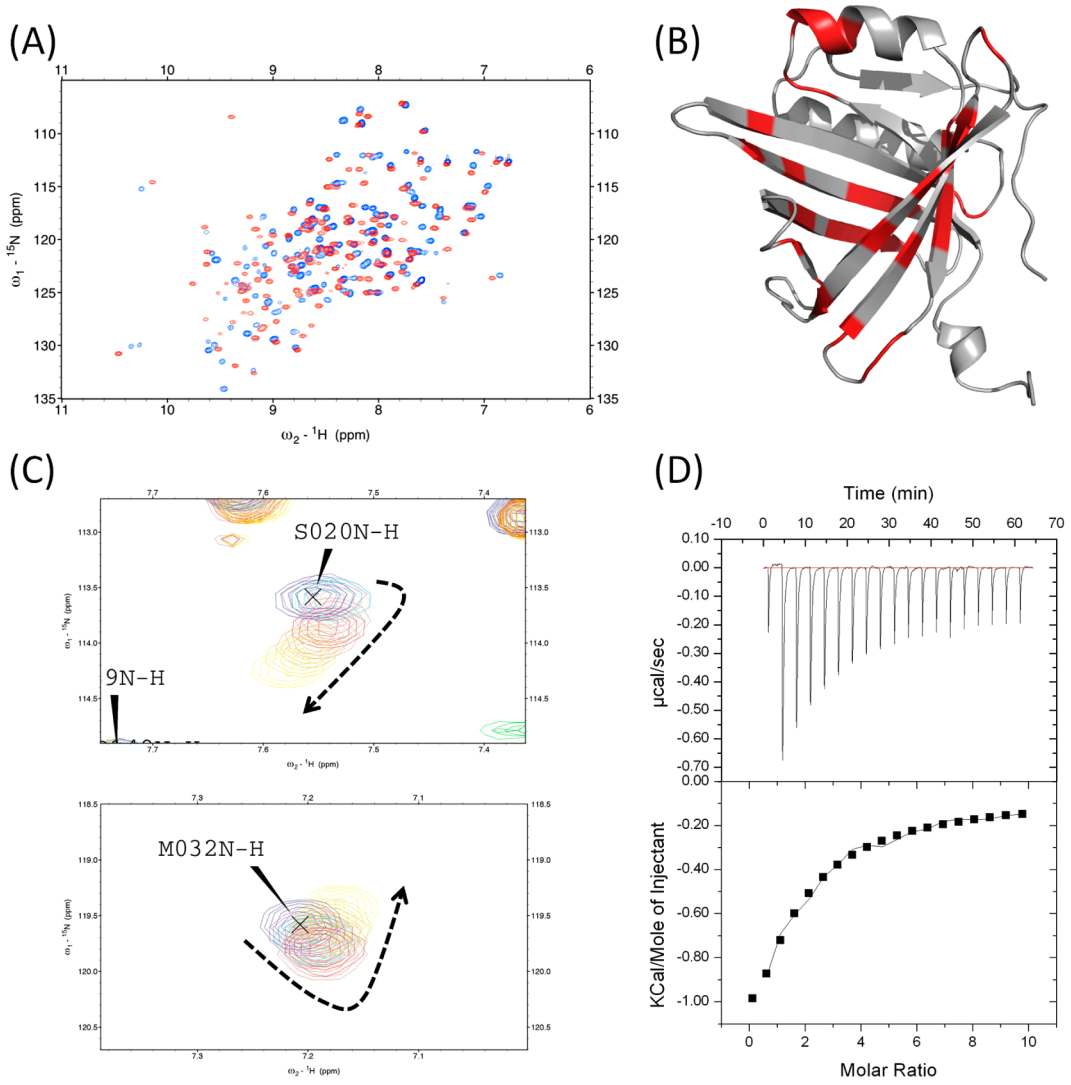


Figure 3. NMR verification of vanillic acid binding to Q83. (A) 2D $^{1}H-^{15}N$ HSQC spectra of unbound Q83 (blue) and bound to vanillic acid (red). (B) Location of observed chemical shift changes mapped onto the 3D solution structure of Q83. Most of the affected residues are located in the calyx, where also the bacterial siderophore enterobactin binds. (C) Nonlinear chemical shift changes are observed as a function of increasing vanillic acid (VA) concentration, indicating the formation of a ternary VA:Q83 (2:1) complex. (D) Binding isotherm of VA probed by isothermal calorimetry (ITC). Raw data (top) and fitted data (bottom) are shown. The experimental ITC data could only be reliably fitted assuming two binding events (K_{D1} = 0.4 mM; K_{D2} = 70 mM). The extracted thermodynamic parameters are given in Table 1.

To illustrate how this novel conceptual approach could be used in drug discovery programs, we further exploited the meta-structure similarity between Q83 and *E. coli* chorismate lyase. As can be seen from Figure 2D, vanillic acid (the ligand of *E. coli* chorismate lyase) shares specific chemical features with the authentic bacterial high-affinity (K_D = 4 nM) ligand of Q83 (the bacterial siderophore enterobactin, Figure 2E), which again underscores the validity of the approach to find suitable ligands just based on primary sequence information. The chemical similarity between vanillic acid and enterobactin is also reflected in the similar binding modes. Figure 2F shows the structural superposition between the solution structure of Q83 bound to enterobactin and *E. coli* chorismate lyase bound to vanillic acid. As can be seen from a close up view (Figure 2G) of the ligand binding sites, both proteins display comparable ligand binding modes. Interestingly, the two vanillic acid molecules that occupy the lyase binding cleft are positioned in an arrangement similar to that of two of the catechol moieties of enterobactin in Q83. We thus concluded that Q83 most likely binds vanillic acid with similar stoichiometry (1:2 molar ratio).

Vanillic acid binding to Q83 was tested by complementary biophysical techniques, such as $^{1}H^{-15}N$ HSQC-based titration experiments and isothermal calorimetry (ITC). Figure 3A shows experimental NMR verification of predicted vanillic acid binding to Q83. While a majority of cross peaks are unaffected by the addition of vanillic acid, there are significant chemical shift changes for a subset of residues lining up in the calyx of Q83 which houses the enterobactin binding site. The NMR data validate the predicted ligand binding and, moreover, provide evidence that the vanillic acid binding site overlaps with the enterobactin binding site (Figure 3B). Interestingly, the nonlinear trajectory of the chemical shift changes as a function of ligand concentration also indicates a more intricate binding mode, presumably due to the formation of a ternary complex with two ligand molecules bound (Figure 3C). Vanillic acid binding was independently verified using ITC (Figure 3D). The data were best fitted using a sequential binding model,

dx.doi.org/10.1021/jm301016m | J. Med. Chem. 2012, 55, 7909−7919

showing that vanillic acid binds to Q83 at a 2:1 molar ratio. The first vanillic acid molecule binds with a K_D of 0.5 mM, whereas the second one binds with a K_D of 70 mM. The obtained thermodynamic parameters (ΔG, ΔH, and ΔS) are given in Table 1. Again, these findings convincingly corroborate the 2:1

Table 1. Thermodynamic Parameters Obtained for Vanillic Acid Binding to Q83

ΔG (kcal/mol)	ΔH (kcal/mol)*	ΔS (cal/mol/deg)
−4.57[a]	−3.1	4.93
−1.47[b]	−56.6	−185

[a]First binding site/event. [b]Second binding site/event.

binding mode of vanillic acid that was already deduced from the 3D structure of the Q83 meta-structure homologue chorismate lyase.

This information about the vanillic acid binding mode and stoichiometry was subsequently used to rationally improve the ligand by a strategy analogous to fragment merging. On the basis of the extracted data, we concluded that merging two vanillic acid fragments should considerably increase the affinity to Q83. The design strategy for improving the ligand capitalized on the available 3D information from the chorismate lyase complex structure. Of particular importance was the location of the functional groups relevant for the interaction with the protein and the intermolecular distances between the two vanillic acid molecules, which provide crucial information about the linker length required to merge the two fragments. To avoid expensive and time-consuming organic chemistry, we pursued an analogue-by-catalogue strategy. As straightforward selection criteria, we used availability of chemically similar compounds (analogue-by-catalogue) from commercial suppliers and good solubility (e.g., sulfonic acid vs carbonic acid group). On the basis of these criteria, we selected 4-amino-1,1′-azobenzene-3,4′-disulfonic acid as an example for an improved hit resulting from a fragment merging step. Binding of 4-amino-1,1′-azobenzene-3,4′-disulfonic acid to Q83 was first monitored by $^{1}H-^{15}N$ HSQC-based titration. Once again, most of the chemical shift changes are located in the calyx (Figure 4A), illustrating that the binding site for 4-amino-1,1′-azobenzene-3,4′-disulfonic acid overlaps with the binding site for vanillic acid and enterobactin. Additionally, the affinity of Q83 for 4-amino-1,1′-azobenzene-3,4′-disulfonic acid was measured by tryptophan fluorescence quenching (Figure 4B). Most importantly, and as expected from the fragment merging strategy, the dissociation constant was found to be substantially lower (26 μM), which fits surprisingly well with the expected affinity values that would be obtained by fusing two vanillic acid molecule ($K_D \sim K_{D1*}K_{D2}$ = 35 μM), the slight difference presumably being due to the slightly modified ligand moiety in 4-amino-1,1′-azobenzene-3,4′-disulfonic acid compared to the original vanillic acid. This example thus convincingly illustrates how meta-structure analysis can be used to identify valid starting points for fragment library design and ligand development (although chemical functionalities such as sulfonic acid and azo groups have pharmacological properties unfavorable for a potential drug candidate). Most importantly, it demonstrates that the information needed to rationally improve molecular fragments, found in a first iteration of an FBLD program, is eventually solely provided by meta-structural data without the requirement of a highly resolved crystal structure.

β-Catenin. The application to β-catenin serves to illustrate how sophisticated (ligand-based) NMR spectroscopy can be used to delineate details of the binding mode and identify possible ligand sites suitable for chemical derivatization (ligand extension). Additionally, we show that meta-structural alignment data can be used to identify molecular fragments with significant protein binding probabilities. Here, the main focus is the exploration of the chemical subspace relevant for (and accessible to) a given protein target. β-Catenin was chosen as a challenging example because of its high medical and pharmaceutical relevance. Interference with the Wnt signaling pathway is a promising and widely pursued approach to control and inhibit vital cellular processes, such as regulation of cell proliferation, differentiation, and apoptosis. Constitutive activation of the canonical Wnt signaling pathway is involved in the development of various human malignancies, such as colorectal carcinomas, melanomas, and ovarian carcinomas.[15] The nuclear activity of β-catenin is connected to its interaction with different transcription factors triggering transcriptional activation of important target genes such as *c-myc, cyclin D1, COX*$_2$, among others.[16] The up-regulation of β-catenin strongly

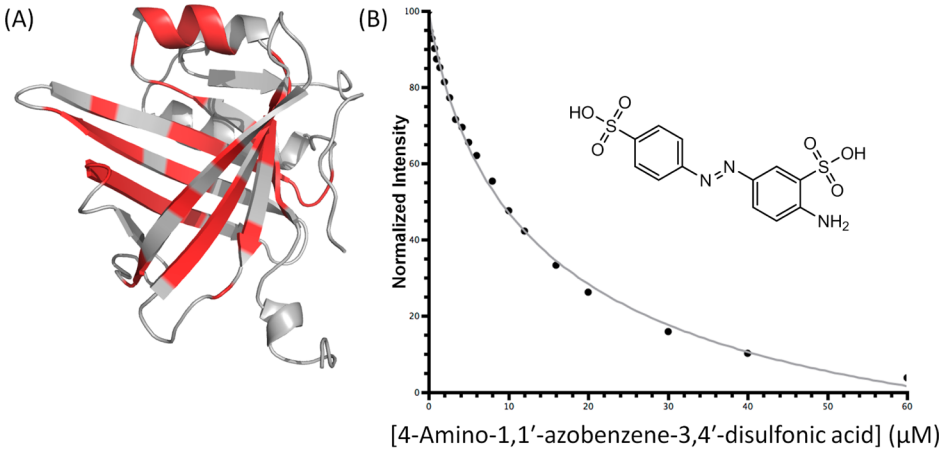


Figure 4. Experimental verification of improved ligand binding to Q83. (A) Location of observed chemical shift changes induced by the improved ligand; 4-amino-1,1′-azobenzene-3,4′-disulfonic acid mapped onto the 3D solution structure of Q83. (B) Fluorescence quench-based measurement of binding affinity. The dissociation constant K_{D*} ($K_{D*} \sim K_{D1*}K_{D2}$) K_{D2} was determined to K_{D*} = 25 μM.

dx.doi.org/10.1021/jm301016m | J. Med. Chem. 2012, 55, 7909−7919

correlates with tumor stage and poor prognosis, and the inhibition of β-catenin activity (e.g., blocking the binding to its authentic protein binding partners) thus offers great potential as an anticancer therapeutic strategy.[17] Given its high medical relevance, small molecules were already developed to inhibit this crucial protein−protein interaction.[18]

The potential of the meta-structure approach to identify possible ligands exclusively based on primary sequence information has already been described.[5] To further evaluate the performance of this approach, we tested whether the already known ligands of β-catenin or chemical fragments thereof are found among the ligands of the best scoring meta-structure homologues. The aim was, of course, not to identify (predict) high-affinity β-catenin binders but to illustrate how the meta-structure approach provides useful information about the accessible chemical space. Overall, the procedure for finding meta-structure homologues was similar to the one described for Q83. The sequence of β-catenin was screened against the protein sequences in the DRUGBANK database and sorted according to meta-structure similarities. A list of best-scoring meta-structure homologues is given in Table 2. The best-scoring hit was trypanothione reductase from *T. cruzi*. In the following we discuss the chemical similarities between the ligands of the identified meta-structure homologues of β-catenin and established β-catenin inhibitors. In the selection of ligands we discarded meta-structure homologues with either peptidic ligands (e.g., hirulog as an established ligand for prothrombin), carbohydrates, or small molecules with very simple chemical functionalities (for example, lactic acid). To have a more representative data set, we also searched in the SuperTarget database[19] for additional ligands reported for the identified meta-structure homologues. A comparison between the chemical structures of the meta-structure-derived ligands together with the structures of the established β-catenin inhibitors is shown in Figure 5. It can clearly be seen that the identified ligands share characteristic chemical moieties with existing β-catenin ligands. For example, the heterocyclic ring of caffeine (a reported ligand of the meta-structure homologue acetylcholine esterase) is found in one of the known β-catenin ligands.[18b] This chemical functionality is also partly embedded in one of the ligands for the best-scoring β-catenin homologue trypanothione reductase.

Table 2. Selection of Best-Scoring Meta-Structure Homologues for Q83 and β-Catenin[a]

Q83	**streptavidin precursor**; chorismate lyase; stromelysin-1; iron-utilization periplasmic protein; FABP; ribosomal small subunit; pseudouridine synthase; phosphatase; androgen receptor
β-catenin	**trypanothione reductase**; malate synthase G; acetohydroxy isomeroreductase; proline dehydrogenase; factor IX; flavocytochrome C; urocanate hydratase; glycodextrin glycosyltransferase; human serum albumin

[a]The best-scoring homologue is given in bold; the meta-structure homologue for which ligand binding was successfully demonstrated is underlined.

Interestingly, fluorescein was identified as a hit based on the meta-structure similarity between β-catenin and human serum albumin. Here again, the observed meta-structure similarity between β-catenin and serum albumin alone does not imply structural identity (or even close homology) but reflects the similar overall α-helical fold of the armadillo-repeat region of human β-catenin and serum albumin. It is important to note that both fluorescein and the chemical analogue eosin Y comprise similar chemical features (e.g., biphenylic ether) to a known β-catenin inhibitor.[18a] In summary, it can be concluded that almost all of the meta-structure-derived ligands comprise to a great extent or in part all the essential ligand functionalities. Thus, they constitute similar chemical scaffolds providing

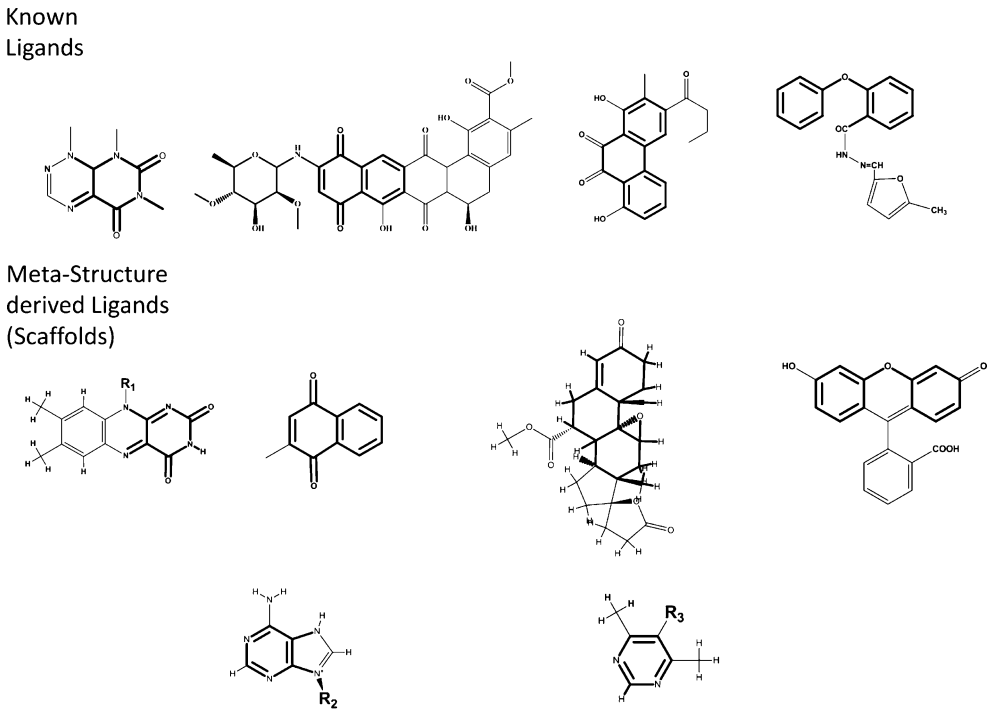


Figure 5. Chemical structures of meta-structure-derived ligands (ligand scaffolds) for β-catenin. The identified ligands using the meta-structure approach (bottom) are compared with known β-catenin inhibitors (top). To illustrate the feasibility of the method to identify relevant chemical fragments, the corresponding moieties of known β-catenin inhibitors are shown in bold.

dx.doi.org/10.1021/jm301016m | J. Med. Chem. 2012, 55, 7909−7919

interaction motifs closely related to the reported ligands for β-catenin.

The unexpected finding of fluorescein and eosin Y being novel ligands for β-catenin was experimentally verified by well-established NMR-STD methodology. Figure 6A,B clearly show that the predicted ligands fluorescein and eosin Y bind to the protein.

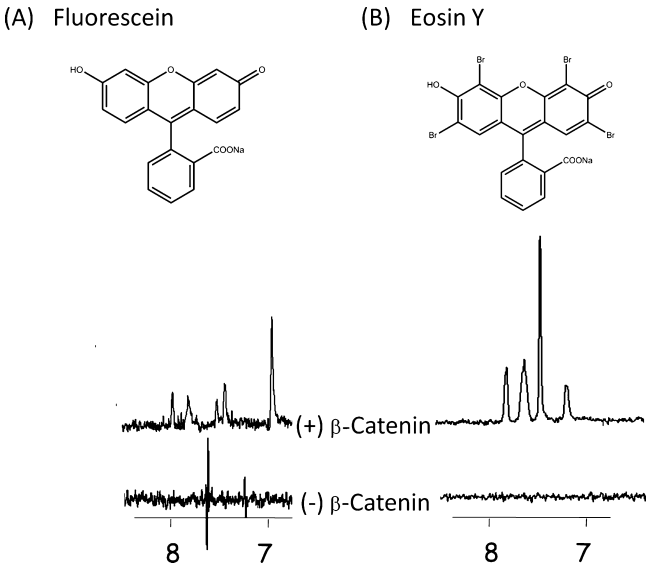


Figure 6. 1D ^{1}H-STD NMR spectra of 30 μM β-catenin in association with 1 mM aqueous solutions of the meta-structure-derived ligands fluorescein and eosin Y. The corresponding negative control in the absence of the protein is shown below each spectrum.

As an illustrative example for how this information could be used for fragment evolution (extension) in a fragment-based lead (drug) design approach, the identified fluorescein and eosin Y ligands were dissected into smaller fragments and tested using information-rich NMR technology. We have recently demonstrated that measuring cross relaxation during adiabatic fast passage provides valuable information about the chemistry of protein−ligand interaction sites.[20] The experimental setup is basically of NOESY type with the conventional mixing time being replaced by an adiabatic RF sweep leading to broad inversion of all protons of the system. By adjusting the RF amplitude of the adiabatic spin lock, a weighted average between ROEs and NOEs can be achieved. Given the dependence (e.g., sign inversion) of the longitudinal cross-relaxation rate on molecular weight, increasing the RF amplitude leads to a sign change of the observed magnetization transfer with a precisely defined zero passage (effective tilt angle with vanishing magnetization transfer). As recently shown, sizable spin diffusion effects, as a result of the existence of dense hydrogen networks or hydrophobic clusters, lead to measurable shifts of the zero passage toward larger tilt angles. Because the magnetization transfer is measured with atomic resolution, this technology allows for ligand-based pharmacophore mapping. Here we applied the technique to probe intraligand magnetization transfer and delineate the binding mode of fluorescein. Figure 7 shows experimental results obtained for fluorescein, indicating that protons located in the biphenylic ether moiety of the ligand display significant spin diffusion effects and are thus in direct contact with the protein and embedded in a hydrophobic binding cleft. We thus concluded that the biphenylic ether system constitutes a relevant (hydrophobic) binding epitope for interaction with β-catenin.

To prove this hypothesis and to delineate the fine details of the interaction pattern between the biphenylic ether moieties found in fluorescein and eosin Y and β-catenin, we studied the interaction between β-catenin and 2-phenoxybenzoate alone (Figure 8A). Positive STD effects observed for 2-phenox-

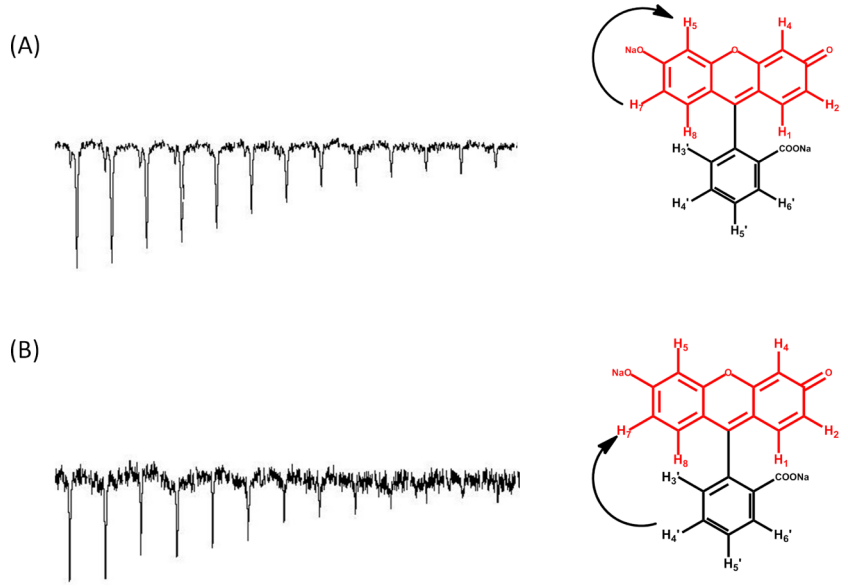


Figure 7. Pharmacophore mapping of the β-catenin fluorescein complex using adiabatic fast passage (AFP) NOESY. Experimental AFP cross-relaxation rates are shown as a function of tilt angle (e.g., increasing RF spin lock amplitude). The following protons were selectively inverted and acted as sources of magnetization transfer: (A) H_1, H_2, H_7, and H_8 from the biphenylic ether fragment and (B) $H_{4'}$, $H_{5'}$, and $H_{6'}$ from the attached aromatic ring system. Protons of H_4 and H_5 of the biphenylic ether scaffold were the detected spins in A while protons of H_1, H_2, H_7, and H_8 were the detected spins in B. The lack of zero passage in both experiments is indicative of the prevalence of spin diffusion effects and thus suggests that the biphenylic ether moiety is embedded in a largely hydrophobic binding cleft.

dx.doi.org/10.1021/jm301016m | J. Med. Chem. 2012, 55, 7909−7919

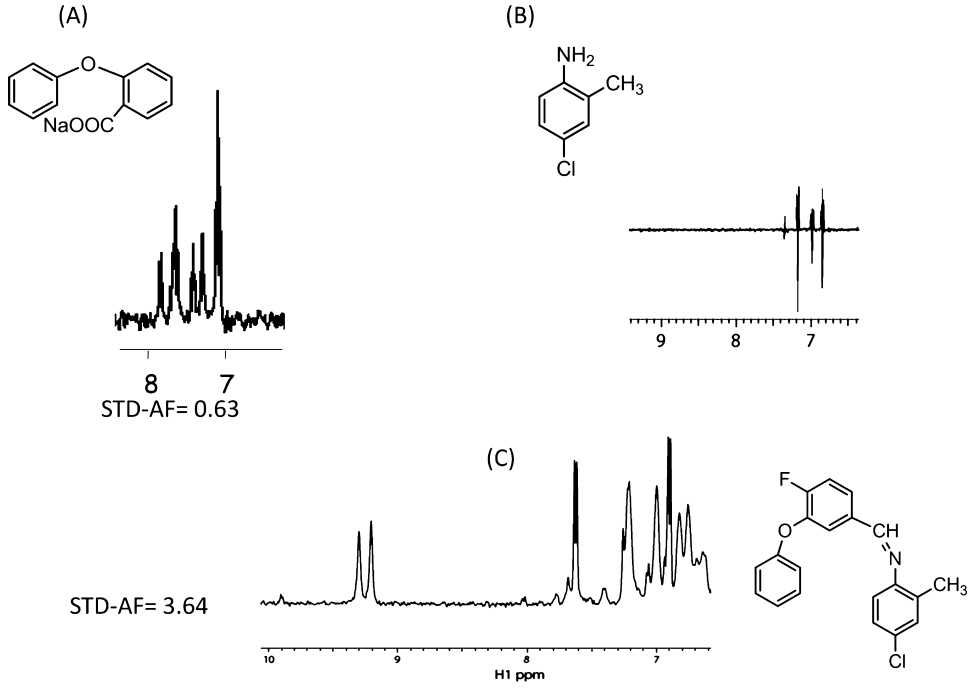


Figure 8. NMR monitoring of fragment extension. 1D ^{1}H-STD NMR is used to observe ligand binding of individual fragments (A, B) and the covalently linked ligand (C). 1D ^{1}H-STD NMR spectra of 30 μM solution of β-catenin in association with 0.5 mM solutions of sodium 2-phenoxybenzoate (A), 4-chloro-2-methylaniline (B), and the merged ligand fused via an imine bond (C). The molecular formulas of the individual compounds are indicated. The experimental data also convincingly confirm the ligand binding deduced by the AFP-NOESY spectrum, as shown in Figure 6. The considerably higher binding affinity of the merged ligand is indicated by the increased average STD amplification factor: 3.64 of (C) vs 0.63 for the sodium 2-phenoxybenzoate (A) at saturation time 1 s.

ybenzoate clearly corroborate the deduced binding mode. It can thus be concluded that the biphenylic ether moiety interacts with a hydrophobic cluster of β-catenin and that the remaining aromatic part in eosin Y/fluorescein is amenable for medicinal chemistry optimization. To evaluate the relevance of these findings for fragment extension and optimization, preliminary (pilot) experiments were performed following a dynamic constitutional combinatorial library strategy originally proposed by Lehn and co-workers.[21] In this approach a dynamic combinatorial library is established via reversible chemical reactions. Here we used an aldehyde analogue of 2-phenoxybenzoate, 4-fluoro-3-phenoxybenzaldehyde, resembling the biphenylic ether moiety and an aromatic amine which can (reversibly) form an imine bond. As can be seen from the results in Figure 8, ligands comprising both functionalities (phenoxybenzaldehyde + amine) have considerably higher relative STD amplification factors (3.64 at saturation time 1 s) in comparison to the single functionality ligand, sodium 2-phenoxybenzoate, which showed lower STD-AF (0.63). On the other hand, the aromatic amino compound alone does not bind to β-catenin. This experimental observation once more corroborates the ligand binding mode deduced from the adiabatic fast passage NOESY experiment and underscores the potential of this novel NMR technology as an information-rich tool for drug development programs. The example described here just serves to illustrate the general strategy. Of course, for realistic ligand optimization, larger fragment (amine) libraries are required. It should also be noted that the chosen ligand (4-fluoro-3-phenoxybenzaldehyde) might be used for ^{19}F-competition screening of large chemical libraries as described by Dalvit et al.[8a] Experiments exploiting these results for the development of novel β-catenin ligands are currently underway in our laboratory and will be published elsewhere.

In summary, the data obtained on β-catenin nicely demonstrates that meta-structural data provide very valuable information about the target relevant chemical subspace. In combination with ligand-based AFP-NOESY NMR spectroscopy, we anticipate fruitful applications to the design of target-optimized fragment libraries with higher enrichment and larger success rates.

■ DISCUSSION AND OUTLOOK

Fragment-based lead discovery (FBLD) is now an accepted strategy for drug discovery in pharmaceutical and biotech companies as well as universities, and more than 10 FBDD-derived leads have already progressed into clinical trials.[22] Despite these tremendous achievements in the past, there are still limitations due to limited protein availability and amenability to current structural biology techniques. Thus, alternative strategies for fragment optimization circumventing the reliance on structure-based approaches are highly needed. Here we demonstrated that it is indeed possible to identify valid chemical starting points for fragment evolution, both fragment merging and extension, without resorting to high-resolution 3D structural information. The approach exploits sequence-derived protein meta-structure information to identify protein homologies unrevealed by conventional (BLAST) sequence alignment strategies, thus significantly broadening the applicability of the approach. Pairwise meta-structure alignments between the protein of interest and templates with experimentally validated small molecule ligands are used to identify chemical scaffolds as possible starting points for lead development. This information can be used to explore the

dx.doi.org/10.1021/jm301016m | J. Med. Chem. 2012, 55, 7909−7919

chemical space accessible by the protein target and provide guidelines for an optimized fragment library design. Additionally, and most importantly, in cases where 3D structures of the identified meta-structure homologues are available, detailed structural information about binding modes can be extracted and exploited for fragment evolution strategies. After experimental verification of ligand binding, information-rich NMR techniques (AFP-NOESY) are applied to provide detailed information about the binding mode that can be used to guide subsequent medicinal chemistry optimization.

This novel approach was demonstrated with applications to two proteins: the quail siderocalin Q83 and the armadillo-repeat region from human β-catenin. The application to quail Q83 illustrates how meta-structure homologies can be used to rationally design strategies to link fragments and improve affinity (fragment merging). The predicted binding mode was successfully verified by NMR spectroscopy. The application to β-catenin demonstrates that protein meta-structure analysis correctly identifies chemical scaffolds (fragments) housing key interaction determinants relevant for ligand binding and thus provides useful information for fragment library design. After experimental verification of binding, AFP-NOESY-NMR spectroscopy was used to probe intraligand magnetization transfer and delineate the relevant binding epitope for interaction with β-catenin.

On the basis of these successful applications and given that only protein sequence information is required, we anticipate widespread applications to hitherto unexplored ('undruggable') protein targets. In particular, as the NMR experiments required for binding mode analysis do not require isotope-labeling and are not limited to low molecular weight macromolecules, challenging targets such as G-protein-coupled receptors or ion channels are amenable. Other 'undruggable' proteins that may be addressed by this approach are intrinsically disordered proteins (IDPs). IDPs are attracting increasing attention in the pharmaceutical sector due to their involvement in fundamental biological processes such as signal transduction, transcriptional control, and protein recruiting to intricate cellular networks. However, the lack of available structural information currently impedes rational drug discovery strategies. Targeting this protein subset via a rational methodology will undoubtedly offer great pharmaceutical potential with beneficial prospects to maintain a target-rich development pipeline.

MATERIALS AND METHODS

Expression/Purification of Recombinant Proteins. Quail recombinant lipocalin Q83 was expressed and purified as described.[13,23] The GST-β-catenin armadillo repeat region was expressed and purified as described by Baminger et al.[24] NMR samples were prepared as follows: (i) Q83 was concentrated up to 1.0 mM in 20 mM NaPi, 50 mM NaCl, 0.5 mM DTT, pH 6.5; (ii) β-catenin was concentrated up to 30 μM in 100 mM Tris, 150 mM NaCl, 2 mM DTT, pH 7.4.

Ligand Preparation. Ligands were purchased from Sigma Aldrich, and stock solutions were prepared in the same buffer as their protein partners. Non-water-soluble ones were dissolved in DMSO-d_6.

NMR Spectroscopy. NMR experiments were carried out at 25 °C on Varian Inova or Direct Drive spectrometers operating at 500, 600, or 800 MHz. All spectra were processed using NMRPipe/NMRDraw[25] and analyzed with Sparky[26] and CARA.[27]

STD-NMR. All STD-NMR spectra were acquired at 500 or 800 MHz. Selective saturation was performed using cascades of Gaussian pulses with a length of 4 ms. On-resonance saturation was applied at −1 ppm and off-resonance saturation at 100 ppm. To eliminate protein resonances from the spectrum, a spin-lock filter (T1ρ-filter) was used. Data were collected at concentrations of 1 mM for fluorescein sodium and eosin Y ligands, and 30 μM protein. Saturation time of 1 s and a ligand concentration of 0.5 mM were used to obtain the STD-AF. STD amplification factor values (STD-AF)[28] were calculated from the following equation:

$$\text{STD} - \text{AF} = \varepsilon(I_0 - I_{\text{sat}})/I_0$$

with ε being the ratio between ligand and protein concentration, $I_0 - I_{\text{sat}}$ the normalized signal intensity of the ligands' protons upon saturation of the protein, and I_0 the signal intensity of the ligand in the reference experiment.

AFP-NOESY. The pulse sequence used is based on a conventional NOESY sequence but has an adiabatic fast passage (AFP) (inversion) pulse during the mixing time.[20] Prior to the adiabatic fast passage, a selective 180° pulse was applied for selective inversion of the source spin (using an IBURP amplitude profile,[29] pulse length of 9.58 ms, and power of 24 db). The experimental details of the AFP were as follows (duration: 400 ms; start of the frequency sweep: 5 kHz downfield of the RF carrier; frequency sweep width: 10 kHz in the upfield direction; RF amplitude: 2 kHz with a 10% sinusoidal and cosinusoidal apodization at the beginning and the end of the sweep). Water suppression was achieved by applying two WATERGATE elements[30] prior to detection. NMR measurements were performed using 30 μM of the human armadillo repeat region of β-catenin and 1 mM fluorescein sodium. Because of very small frequency differences in the fluorescein 1D ^{1}H spectrum, two sets of fluorescein protons were inverted in two independent experiments: (A) H_1, H_2, H_7, and H_8 and (B) $H_{4'}$, $H_{5'}$, and $H_{6'}$; intraligand NOEs to H_4 and H_5 (in experiment A) and H_1, H_2, H_7, and H_8 (in experiment B) were monitored.

Fluorescence Quenching Binding Assay. Fluorescence quenching of lipocalin Q83 was measured on a Perkin-Elmer LS 50B fluorimeter with 5 nm slit band-pass, using the characteristic excitation and emission wavelengths λ_{exc} = 280 nm and λ_{em} =340 nm. Measurements were made at a protein concentration of 2 μM in 20 mM NaPi, 50 mM NaCl, 0.5 mM DTT, pH 6.5 at 25 °C. The volume of the cell was 2 mL. The decrease of fluorescence intensity was followed upon addition of a concentrated ligand solution (200 μM). The decrease of fluorescence intensity was plotted as a function of the ligand concentration. Experimental data points were fitted using QtiPlot, assuming a single binding site model, and with [P], [L], K_D, I_{max}, and I_{sat} being the protein concentration, ligand concentration, dissociation constant, reference intensity, and intensity at saturating concentration of the ligand, respectively.

Isothermal Calorimetry. Calorimetric measurements were carried out at 25 °C using a μ-ITC calorimeter (MicroCal). Samples were extensively dialyzed against 20 mM NaPi, 50 mM NaCl, pH 6.5. The protein concentration in the cell was 1 mM, the protein was titrated by 20 successive injections of vanillic acid at 20 mM. Heats of dilution were measured in blank titration by injecting the ligand into the buffer and subtracted from the binding heat. Thermodynamic parameters were determined by nonlinear least-squares method using routines included in the Origin software package (MicroCal).

DRUGBANK Screening. The screening of targets from the DRUGBANK database using meta-structure-based sequence alignment was implemented as described previously.[5] In a first step, the meta-structure parameters of DRUGBANK target sequences were calculated. Second, instead of using well-established measures for amino acid similarities such as BLOSUM62, calculated meta-structure parameters derived from the primary sequence were used to define pairwise similarity matrices. Details of the scoring function using both second structure and compactness values can be found elsewhere.[5] A global analysis using second structure and compactness distributions were used to (pre)filter the target sequences. On average, only about 3500 target sequences were subjected to the calculation of sequence alignments. The hits are scored according to the similarity measures. Typically, only hits with a relative similarity score larger than 0.60 (normalized to the maximum score) are considered.

The strategy is based on the consideration that meta-structure similarities in ligand interaction sites provide valuable starting information for the identification of chemical scaffolds and guiding

structures in ligand development programs without the requirement of high-resolution protein structures. Suitable protein target sequences for meta-structure alignment can be taken from, for example, the DRUGBANK database, a public repository of biologically relevant protein targets with experimentally verified inhibitory ligands.[12]

AUTHOR INFORMATION

Corresponding Author

*Phone: ++43-1-4277-52202; e-mail: robert.konrat@univie.ac.at.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by grants from the Austrian Science Foundation FWF P20549, P22125, and W1221-B03, and ÖAD (Ku mit Agypten, art 9/2). The authors thank Georg Kontaxis (University of Vienna) for NMR support.

ABBREVIATIONS USED

AFP, adiabatic fast passage; AFP-NOESY, adiabatic fast passage nuclear Overhauser effect spectroscopy; BLAST, basic local alignment search tool; FABP, fatty acid binding protein; FAXS, fluorine chemical shift anisotropy and exchange for screening; FBLD, fragment-based lead discovery; HSQC, heteronuclear single-quantum coherence; HTS, high-throughput screen; IDP, intrinsically disordered proteins; ITC, isothermal calorimetry; NOE, nuclear Overhauser effect; PMSSC, protein meta-structure similarity clustering; PSSC, protein structure similarity clustering; ROE, rotating frame Overhauser effect; SCN, sideroclain; STD, saturation transfer difference; STD-AF, saturation transfer difference amplification factor

REFERENCES

(1) Salum, L. B.; Andricopulo, A. D. Fragment-based QSAR: perspectives in drug design. *Mol. Diversity* **2009**, *13*, 277−285.

(2) Foloppe, N. The benefits of constructing leads from fragment hits. *Future Med. Chem.* **2011**, *3*, 1111−1115.

(3) Zhou, J. Z. Structure-directed combinatorial library design. *Curr. Opin. Chem. Biol.* **2008**, *12*, 379−385.

(4) Schulz, M. N.; Hubbard, R. E. Recent progress in fragment-based lead discovery. *Curr. Opin. Pharmacol.* **2009**, *9*, 615−621.

(5) Konrat, R. The protein meta-structure: a novel concept for chemical and molecular biology. *CMLS, Cell. Mol. Life Sci.* **2009**, *66*, 3625−3639.

(6) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403−410.

(7) (a) Mayer, M.; Meyer, B. Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angew. Chem., Int. Ed.* **1999**, *38*, 1784−1788. (b) Chen, A.; Shapiro, M. J. NOE pumping: A novel NMR technique for identification of compounds with binding affinity to macromolecules. *J. Am. Chem. Soc.* **1998**, *120*, 10258−10259. (c) Dalvit, C.; Pevarello, P.; Tato, M.; Veronesi, M.; Vulpetti, A.; Sundstrom, M. Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. *J. Biomol. NMR* **2000**, *18*, 65−68.

(8) (a) Dalvit, C. Ligand- and substrate-based F-19 NMR screening: Principles and applications to drug discovery. *Prog. Nucl. Magn. Reson. Spectrosc.* **2007**, *51*, 243−271. (b) Dalvit, C.; Fagerness, P. E.; Hadden, D. T. A.; Sarver, R. W.; Stockman, B. J. Fluorine-NMR experiments for high-throughput screening: Theoretical aspects, practical considerations, and range of applicability. *J. Am. Chem. Soc.* **2003**, *125*, 7696−7703.

(9) Sanchez-Pedregal, V. M.; Reese, M.; Meiler, J.; Blommers, M. J. J.; Griesinger, C.; Carlomagno, T. The INPHARMA method: Protein-mediated interligand NOEs for pharmacophore mapping. *Angew. Chem., Int. Ed.* **2005**, *44*, 4172−4175.

(10) (a) Ludwig, C.; Michiels, P. J. A.; Wu, X.; Kavanagh, K. L.; Pilka, E.; Jansson, A.; Oppermann, U.; Gunther, U. L. SALMON: Solvent accessibility, ligand binding, and mapping of ligand orientation by NMR Spectroscopy. *J. Med. Chem.* **2008**, *51*, 1−3. (b) Ludwig, C.; Michiels, P. J. A.; Lodi, A.; Ride, J.; Bunce, C.; Gunther, U. L. Evaluation of solvent accessibility epitopes for different dehydrogenase inhibitors. *ChemMedChem* **2008**, *3*, 1371−1376.

(11) Dekker, F. J.; Koch, M. A.; Waldmann, H. Protein structure similarity clustering (PSSC) and natural product structure as inspiration sources for drug development and chemical genomics. *Curr. Opin. Chem. Biol.* **2005**, *9*, 232−239.

(12) Wishart, D. S.; Knox, C.; Guo, A. C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J., DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **2006**, *34* (Database issue), D668−D672.

(13) Coudevylle, N.; Geist, L.; Hotzinger, M.; Hartl, M.; Kontaxis, G.; Bister, K.; Konrat, R. The v-myc-induced Q83 lipocalin is a siderocalin. *J. Biol. Chem.* **2010**, *285*, 41646−41652.

(14) (a) Bachman, M. A.; Miller, V. L.; Weiser, J. N. Mucosal lipocalin 2 has pro-inflammatory and iron-sequestering effects in response to bacterial enterobactin. *PLoS Pathog.* **2009**, *5*, e1000622. (b) Yang, J.; Goetz, D.; Li, J. Y.; Wang, W.; Mori, K.; Setlik, D.; Du, T.; Erdjument-Bromage, H.; Tempst, P.; Strong, R.; Barasch, J. An iron delivery pathway mediated by a lipocalin. *Mol. Cell* **2002**, *10*, 1045−1056. (c) Leng, X.; Lin, H.; Ding, T.; Wang, Y.; Wu, Y.; Klumpp, S.; Sun, T.; Zhou, Y.; Monaco, P.; Belmont, J.; Aderem, A.; Akira, S.; Strong, R.; Arlinghaus, R. Lipocalin 2 is required for BCR-ABL-induced tumorigenesis. *Oncogene* **2008**, *27*, 6110−6119. (d) Shi, H.; Gu, Y.; Yang, J.; Xu, L.; Mi, W.; Yu, W. Lipocalin 2 promotes lung metastasis of murine breast cancer cells. *J. Exp. Clin. Cancer Res.* **2008**, *27*, 83.

(15) Kobayashi, M.; Honma, T.; Matsuda, Y.; Suzuki, Y.; Narisawa, R.; Ajioka, Y.; Asakura, H. Nuclear translocation of beta-catenin in colorectal cancer. *Br. J. Cancer* **2000**, *82*, 1689−1693.

(16) Camilli, T. C.; Weeraratna, A. T. Striking the target in Wnt-y conditions: intervening in Wnt signaling during cancer progression. *Biochem. Pharmacol.* **2010**, *80*, 702−711.

(17) Ovaa, H.; Kuijl, C.; Neefjes, J. Recent and new targets for small molecule anti-cancer agents. *Drug Discovery Today: Technol.* **2009**, *6*, e3−e11.

(18) (a) Trosset, J. Y.; Dalvit, C.; Knapp, S.; Fasolini, M.; Veronesi, M.; Mantegani, S.; Gianellini, L. M.; Catana, C.; Sundstrom, M.; Stouten, P. F.; Moll, J. K. Inhibition of protein-protein interactions: the discovery of druglike beta-catenin inhibitors by combining virtual and biophysical screening. *Proteins* **2006**, *64*, 60−67. (b) Lepourcelet, M.; Chen, Y. N.; France, D. S.; Wang, H.; Crews, P.; Petersen, F.; Bruseo, C.; Wood, A. W.; Shivdasani, R. A. Small-molecule antagonists of the oncogenic Tcf/beta-catenin protein complex. *Cancer Cell* **2004**, *5*, 91−102.

(19) Gunther, S.; Kuhn, M.; Dunkel, M.; Campillos, M.; Senger, C.; Petsalaki, E.; Ahmed, J.; Urdiales, E. G.; Gewiess, A.; Jensen, L. J.; Schneider, R.; Skoblo, R.; Russell, R. B.; Bourne, P. E.; Bork, P.; Preissner, R. SuperTarget and Matador: resources for exploring drug-target relationships. *Nucleic Acids Res.* **2008**, *36* (Database issue), D919−D922.

(20) Auer, R.; Kloiber, K.; Vavrinska, A.; Geist, L.; Coudevylle, N.; Konrat, R. Pharmacophore mapping via cross-relaxation during adiabatic fast passage. *J. Am. Chem. Soc.* **2010**, *132*, 1480−1481.

(21) Ramstrom, O.; Lehn, J. M. Drug discovery by dynamic combinatorial libraries. Nature reviews. *Drug Discovery* **2002**, *1*, 26−36.

(22) (a) Congreve, M.; Chessari, G.; Tisi, D.; Woodhead, A. J. Recent developments in fragment-based drug discovery. *J. Med. Chem.* **2008**, *51*, 3661−3680. (b) Hajduk, P. J.; Greer, J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat. Rev. Drug Discovery* **2007**, *6*, 211−219.

(23) Hartl, M.; Matt, T.; Schuler, W.; Siemeister, G.; Kontaxis, G.; Kloiber, K.; Konrat, R.; Bister, K. Cell transformation by the v-myc

oncogene abrogates c-Myc/Max-mediated suppression of a C/EBP beta-dependent lipocalin gene. *J. Mol. Biol.* **2003**, *333*, 33−46.
(24) Baminger, B.; Ludwiczek, M. L.; Kontaxis, G.; Knapp, S.; Konrat, R. Protein-protein interaction site mapping using NMR-detected mutational scanning. *J. Biomol. NMR* **2007**, *38*, 133−137.
(25) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **1995**, *6*, 277−293.
(26) Goddard, T. D.; Kneller, D. G. *Sparky 3*, 2002.
(27) Keller, R., *The computer aided resonance assignment tutorial*; Cantina Verlag: Goldau, Switzerland, 2004.
(28) Angulo, J.; Enriquez-Navas, P. M.; Nieto, P. M. Ligand-receptor binding affinities from saturation transfer difference (STD) NMR spectroscopy: the binding isotherm of STD initial growth rates. *Chemistry* **2010**, *16*, 7803−7812.
(29) Geen, H. F., R. Band-selective radiofrequency pulses. *J. Magn. Reson.* **1991**, *93*, 93−141.
(30) Piotto, M.; Saudek, V.; Sklenar, V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* **1992**, *2*, 661−665.
(31) Sippl, M. J.; Wiederstein, M. A note on difficult structure alignment problems. *Bioinformatics* **2008**, *24*, 426−427.
(32) The PyMOL Molecular Graphics System, V. r. p., Schrödinger, LLC.