Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters 15 (2005) 1779–1783
Bioorganic & Medicinal Chemistry Letters

QSAR-by-NMR: quantitative insights into structural determinants for binding affinity by analysis of $^{1}H/^{15}N$ chemical shift differences in MMP-3 ligands

Hans Matter,[a,*] Manfred Schudok,[a] Bettina Elshorst,[b] Doris M. Jacobs,[b] Krishna Saxena[b] and Herbert Kogler[a]

[a]*Aventis Pharma Deutschland GmbH, DI&A Chemistry, Building G 878, D-65926 Frankfurt am Main, Germany*
[b]*Johann Wolfgang Goethe-Universität, Institut für Organische Chemie, Marie-Curie Str. 11, D-60439 Frankfurt am Main, Germany*

Received 6 October 2004; revised 3 February 2005; accepted 17 February 2005

Abstract—A novel strategy is applied to obtain quantitative insights on factors influencing biological affinity in protein–ligand complexes. This approach is based on the detection of ligand binding by ^{15}N and ^{1}H amide chemical shift differences in two-dimensional ^{15}N-heteronuclear single-quantum correlation spectra. Essential structural features linked to affinity can be extracted using statistical analysis of ^{15}N and ^{1}H amide chemical shift differences in congeneric series relative to uncomplexed protein spectra, as demonstrated for 20 MMP-3 inhibitors in complex with human matrix metalloproteinase stromelysin (MMP-3). The statistical analysis using PLS led to a significant model, while its chemical interpretation, highlighting the importance of particular residues for affinity, are in agreement to an X-ray structure of one key compound in the homologue MMP-8 binding site.

Information about ligand binding modes is of great interest for modern medicinal chemistry to guide lead optimization. To achieve complementarity between ligands and binding sites, an understanding of favorable determinants for protein–ligand interactions is required. Over the past years, NMR-based screening methods have emerged as powerful techniques augmenting high-throughout screening to identify lead structures.[1–3] As NMR can identify true hits interacting with the target, it allows to reject false positives in combination with HTS.[4] While NMR-based methods have mainly been used to identify ligands binding to a receptor, they also found application in pioneering studies toward identification of small molecules binding to distinct protein subsites (secondary site screening[5]), followed by synthesis to link these fragments, which result in potent inhibitors and preliminary insights into structure–activity relationship (SAR by NMR[6,7]).

Based on the principle to detect ligand binding by ^{15}N and ^{1}H amide chemical shift differences in two-dimensional ^{15}N-heteronuclear single-quantum correlation spectra (^{15}N-HSQC[8]), we report a novel strategy to obtain quantitative information on factors influencing affinity in protein–ligand complexes. Essential features linked to biological affinity are extracted using statistical analysis of ^{15}N and ^{1}H amide chemical shift differences in congeneric series relative to uncomplexed protein spectra. In particular we apply partial-least-squares regression (PLS[9]) to correlate chemical shift differences to biochemical binding affinities. As chemical shift perturbations are detected on the protein site, no a priori assumption about ligand binding modes is required to derive a model. However, chemical interpretation is supported by experimental or postulated binding modes.

This approach was successfully applied to 20 human stromelysin (MMP-3) inhibitors[10] selected on the basis of solubility and diversity. Matrix metalloproteinases (MMPs) are a family of zinc endopeptidases, which are capable of degrading the extracellular matrix of connective tissues.[11] They are implicated in degenerative diseases in which there is a slow matrix degradation rate, including cartilage loss in osteoarthritis[12] and rheumatoid arthritis,[13] bone matrix degradation in osteoporosis or remodeling in Alzheimer disease.[14] All MMPs possess three domains: a propeptide cleaved during activation, a

Keywords: QSAR; NMR; MMP-3; Protein–ligand interactions.
* Corresponding author. Tel.: +49 69 305 84329; fax: +49 69 331399; e-mail: hans.matter@aventis.com

0960-894X/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2005.02.048

catalytic domain (~180 residues) with a conserved **HEXXHXXGXXH** zinc binding motif, and a hemopexin-like domain. Stromelysin (MMP-3), the intestinal fibroblast collagenase (MMP-1), and the neutrophil collagenase (MMP-8) are responsible for the cleavage of type I-, II-, and III-collagen, and are regarded as key enzymes in the pathology of matrix degradation. The 1,2,3,4-tetrahydroisoquinoline scaffold provides an ideal geometry to link hydroxamates and carboxylates as zinc-complexing groups with S1′-pocket-directed[15] substituents for this hydrophobic subsite.

Chemical shift perturbation can be efficiently used to determine binding constants (K_D) in the case of fast exchange.[16] As the chemical shift in the pure bound state cannot be experimentally determined for weak binders, a fitting procedure is required to account for this parameter.[17] Hence, we used a series of medium to high-affinity ligands, where slow exchange resulted in two sets of signals at sub-stoichiometric ligand concentrations, which allows to directly extract chemical shift differences between free and bound state without compensating for varying concentrations of co-solvents or pH usually complicating the interpretation of titration series.

The generation of hypotheses on the intrinsic relationship between affinity as response variable and independent descriptors like chemical shift differences plus assessing the statistical significance of any resulting model are the major steps in building quantitative structure–activity relationships. For large datasets with highly collinear variables, PLS[9] is one method of choice, because conventional multiple linear regression (MLR) only deals with situations, in which the number of rows is at least three times larger than the number of variables. Furthermore, the individual variables for MLR are assumed to be independent and uncorrelated. PLS is a regression method in analogy to a principal component analysis (PCA) based on latent, orthogonal variables. The directions of these latent variables are slightly shifted from the PCA solution to obtain optimal correlations between the vectors of X- and Y-blocks (activity vs descriptors). PLS thus is suited to derive a linear relationship for underdetermined matrices, while crossvalidation is used to check for consistency and predictivity of the model.

The chemical shift differences of backbone amide $^{1}H/^{15}N$ resonances in ^{15}N-MMP-3 HSQC spectra[18] were extracted for 81 amino acids surrounding the inhibitor binding site in the MMP-3 catalytic domain and referenced to 2D-HSQC spectra of uncomplexed MMP-3 under identical conditions (Fig. 1). This resulted in a data table of 162 chemical shift differences as descriptors in the X-block and 20 biological activities[19] as dependent variables. After PLS analysis and crossvalidation on this input dataset, a significant model with 3 PLS components (i.e., latent orthogonal variables) resulted with 72 descriptors.[20] Crossvalidation indicated the significance and predictivity of this model: a leave-one-out crossvalidated r^2 (q^2) value of 0.657 (SD 0.675) and a conventional r^2 value of 0.976 (SD 0.177) were obtained. Further statistical validation underscored the model's significance to correlate chemical shifts to biological activity, for example, leave-two-out crossvalidation (q^2 0.629) and repeated analyses using five crossvalidation groups (q^2 0.578). To further validate this model, the dataset was split into training and test datasets consisting of 16 and 4 compounds, respectively. For the training dataset a three component PLS model was obtained with 85 descriptors, a q^2 value of 0.471 and a conventional r^2 value of 0.959. Applying this model to the test set revealed a reliable prediction of these test set molecules with a standard error of prediction (SDEP) of 0.637. Hence, this approach correctly predicted binding affinities for these ligands within 0.6 orders of magnitude.

PLS does not only produce a quantitative model for affinity prediction, but it is also useful to highlight the importance of particular variables to enhance affinity. Descriptors with positive PLS coefficients (Fig. 2A and Supporting information) have a positive correlation to biological affinity. In other words, a more positive chemical shift difference for this particular residue is related to an increase of binding affinity (i.e., lower IC_{50} value). In contrast, negative PLS coefficients indicate that either positive chemical shift differences decrease affinity or negative chemical shift differences increase affinity. Hence, a combined interpretation in terms of relevant amino acid backbone amides interacting with potential ligands becomes possible from the perspective of the protein binding site.

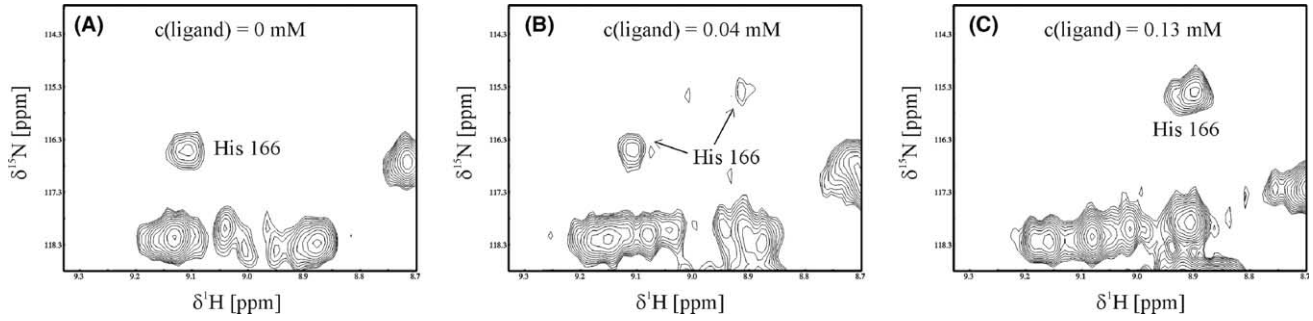

Figure 1. Selected regions from 500 MHz $^{1}H/15N$ HSQC-spectra for uncomplexed MMP-3 (A) and its complex with inhibitor **62** (IC_{50} 500 nM) in different concentrations (B, C). As high-affinity results in slow exchange, two sets of signals are observed for uncomplexed and complexed MMP-3, as exemplified for His166.

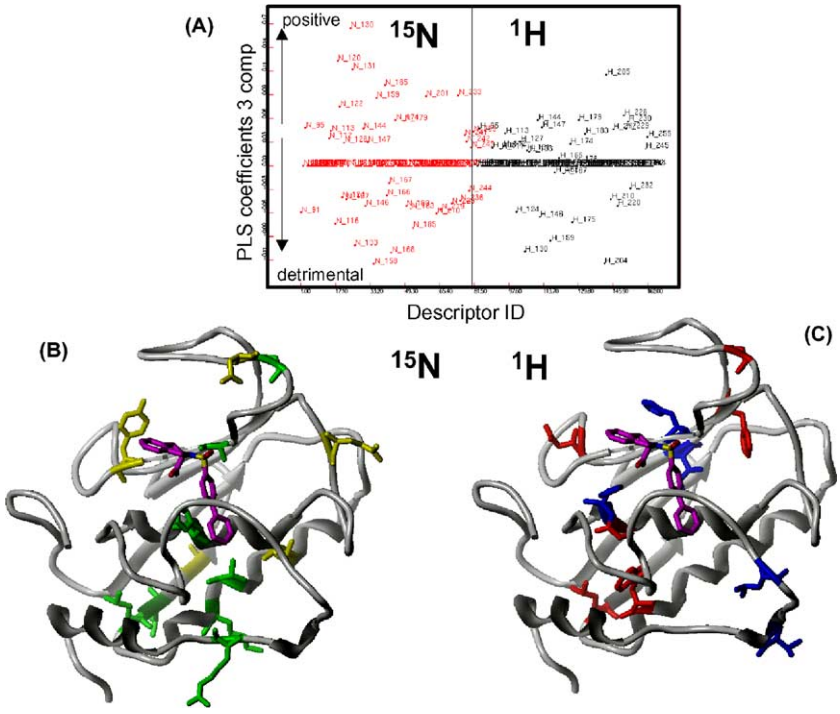

Figure 2. (A) PLS coefficient plot indicating the relative importance of shift differences for ^{15}N (red) and ^{1}H (black). (B) Mapping of important PLS coefficients for ^{15}N onto the MMP-3 structure from PDB entry 1caq with **62** docked into the active site (magenta). Green residues (coeff > 0.07) indicate more positive ^{15}N shift differences to increase affinity. Yellow residues (coeff < 0.07) indicate more negative ^{15}N shift difference to increase affinity. (C) Similar analysis for ^{1}H. Blue residues (>0.05) indicate more positive ^{1}H shift differences to increase affinity. Red residues indicate more negative ^{1}H shift differences to increase affinity.

Relevant ^{1}H and ^{15}N chemical shift differences from the PLS coefficient plot (Fig. 2A) were mapped onto the corresponding amino acids in the MMP-3 protein structure[21] for interpretation. These PLS coefficients (Supporting information) indicate the relative importance of chemical shifts to explain biological affinity. Green residues in Figure 2B indicate that a positive ^{15}N shift difference for the backbone amide nitrogen increases biological affinity, while a more negative shift is detrimental. Yellow amino acids indicate more negative ^{15}N shift difference to increase affinity, while positive shifts are detrimental. A similar analysis is possible for ^{1}H chemical shift differences, as indicated in Figure 2C. Blue indicates that a more positive ^{1}H shift difference increases affinity, while red residues highlight amide protons, where negative ^{1}H shift differences increases affinity and vice versa.

As chemical shift differences upon binding of a single ligand are difficult to explain in structural terms, we present a comparative analysis for ligands with different affinities, allowing for the identification of essential binding site regions for ligand interactions. Amide NH groups serve as a indirect probe to indicate structural changes in their neighborhood, which could be caused by chemical shielding upon ligand binding, side chain movements or altered hydrogen bond geometries.

This quantitative model from chemical shift differences is only based on experimental data from NMR and biochemical assays. Its interpretation is then guided by experimental or postulated binding poses. The combination of both approaches allows to highlight key regions in the binding site involved in protein–ligand recognition, as indicated by significant shift differences (Fig. 3). Hence, an interpretation in structural terms is given below using the experimental binding mode of **62** as reference structure in the homologous human neutrophil collagenase (MMP-8) at 1.7 Å resolution.[22] The tetrahydroisoquinoline-3-carboxylate **62** with a biphenyl-sulfonamide moiety directed toward S1′ shows IC_{50} values of 500 nM for MMP-3 and 10 nM for MMP-8, respectively.

The positive ^{15}N shift contribution for Ala120**N** can be attributed to bulky substituents in S1′, which influence Phe196 and Leu197 side chain conformations next to Ala120. Replacing the distal phenyl ring in **62** by Br results in a dramatic loss of affinity (**90**: 800,000 nM), while the addition of a 4-F or 4-Cl substituent increases complementarity and thus binding affinity (**71**: 400 nM; **49**: 100 nM). Further orienting the biphenyl group into S1′ by inserting an ethyl-spacer results in high-affinity binding (**86**: 50 nM). These changes in S1′ consistently affect neighboring amino acids, leading to the correlation between chemical shifts and MMP-3 affinity. Filling S1′ by appropriate substituents is known to affect affinity.[10]

The influence for Ala165**N**, involved in a hydrogen bond to an inhibitor sulfonamide oxygen, is due to changes in hydrogen-bond geometry or conformations of neighboring residues caused by ligands, which fit less

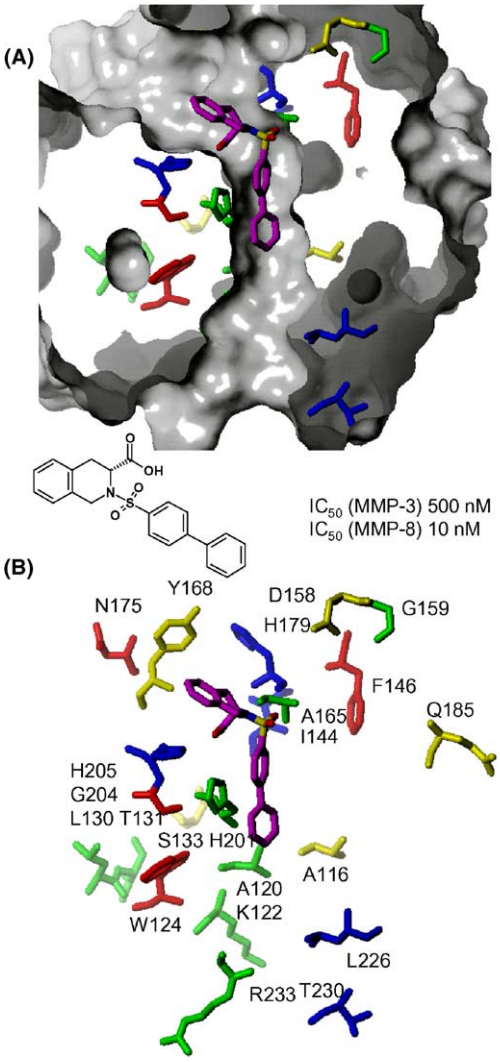

Figure 3. (A) Binding mode of **62** (magenta) in MMP-3 (1caq) from docking with key amino acids from the three component PLS model. The binding site is indicated by a solvent accessible surface, the color code is according to Figure 2 with green and yellow indicating important ^{15}N shift differences and blue and red for ^{1}H shift differences, respectively. (B) Similar view without protein surface.

optimal into S1′. Another influence is obvious for His201**N**, a residue involved in zinc binding and interactions to an inhibitor aromatic ring in S1′. Conformational changes or modulated electrostatics of the imidazole ring affect the corresponding ^{15}N shift. Those changes can be attributed to different zinc-binding groups. While a hydroxamate with a small S1′ directed substituent leads to high affinity, the replacement against a carboxylic acid reduces this affinity significantly (**6**: 20 nM vs **5** 10,000 nM). This loss is only compensated for by larger S1′ substituents, as already discussed. Another positive influence is observed for the flexible Arg233 at the bottom of S1′, which again highlights the importance of conformational changes caused by appropriate inhibitor size and properties in S1′. Replacing the biphenyl substituent in **62** by a biarylether results in an increase in affinity (**52**: 200 nM), while any additional substituent in 4-position now is no longer directed toward the flexible amino acids at the bottom of S1′, consequently leading to a loss of affinity (**84** (–CN): 1000 nM).

An indirect effect is observed for Leu130 and Thr131 amide N, which are away from the binding site, but significantly contribute to the PLS coefficients. For distant side chains the interpretation is less obvious, as experimental differences might result from long-distance 'relay' effects to amide shifts. Our analysis shows that these influences indeed follow significant trends. Negative ^{15}N shift contribution are observed for Asp158 on top of the Val163 side chain, which itself forms the entrance into the hydrophobic S1′ pocket. Different inhibitors might affect Val163 and thus Asp159 indirectly. Further evidence for the importance of this region is provided by Gly159**N** and **H** PLS coefficients. The ^{15}N chemical shift for Tyr168 is influenced by the conformation of three aromatic rings from His166, Tyr155, and Tyr168, which could be directly modified upon ligand binding.

The positive ^{1}H shift coefficient for His179 could be explained by its hydrogen bond to the carbonyl oxygen of His166, as this latter residue is directly involved in ligand recognition. For His205, involved in zinc-binding, ligand interaction to neighboring aromatic rings of His201/205 and its involvement in a hydrogen bond with His201 are reflected by PLS coefficients. Negative ^{1}H shift contributions are observed for Gly204, which is also close to His201/205 and thus directly affected by possible conformational changes upon ligand binding. It also forms the third partner in a bifurcated hydrogen bond to His166C═O (3.1, 3.3 Å distance). Finally Trp125**H** is significantly affected by size and nature of the inhibitor substituent in S1′. Especially the distal pyridyl-ring in **24** (20 nM) causes a negative ^{1}H shift, which might be due to a significant change in electrostatic potential in this pocket.

Hence, we have established a correlation between chemical shift differences of a series of MMP-3 ligands and their biological affinity. Binding of ligands to a protein introduces changes in the binding site at distinct positions, depending on the degree of steric, and electrostatic complementarity. Those changes are translated in a consistent way into amino acid chemical shifts. Subtle differences in ligand substitutions and binding modes in this series are responsible for shifts in neighboring residues. NMR provides a rapid experimental probe to detect those changes, and relate them to altered protein–ligand interactions. The fact that affinities for all 20 molecules could be explained by a single model strongly supports the hypothesis of a consistent binding mode for this series.[10,22] The provided interpretation in structural terms provides evidence to this approach and links the statistical model to known structure–activity relationships.

To our knowledge this is the first example of such a strategy combined with an independent experimental approach. This illustrates its potential to obtain relevant information about ligand binding modes and critical binding site regions for this protein. In future applications to other targets, it remains to be unveiled, whether the striking correlation between chemical shift differences

and biological activity is a general trend. We suggest to use the term 'QSAR-by-NMR' to describe quantitative models for correlating chemical shift differences to binding affinities. This method in combination with an interpretation in structural terms provides an understanding of protein–ligand interaction features in structure-driven drug discovery.

Supplementary data

Two tables with structures and affinities of MMP-3 inhibitors and coefficients for the final PLS model. Supplementary data associated with this article can be found, in the online version at doi:10.1016/j.bmcl.2005.02.048.

References and notes

1. Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. *Q. Rev. Biophys.* **1999**, *32*, 211–240.
2. Coles, M.; Heller, M.; Kessler, H. *Drug Discovery Today* **2003**, *8*, 803–810.
3. Pellecchia, M.; Sem, D. S.; Wüthrich, K. *Nat. Rev. Drug Disc.* **2002**, *1*, 211–219.
4. Van Dongen, M. J. P.; Uppenberg, J.; Svensson, S.; Lundbäck, T.; Akerud, T.; Wikström, M.; Schultz, J. *J. Am. Chem. Soc.* **2002**, *124*, 11874–11880.
5. Jahnke, W.; Perez, L. B.; Paris, C. G.; Strauss, A.; Fendrich, G.; Nalin, C. M. *J. Am. Chem. Soc.* **2000**, *122*, 7394–7395.
6. Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. *Science* **1996**, *274*, 1531–1534.
7. Hajduk, P. J.; Sheppard, G.; Nettesheim, D. G.; Olejniczak, E. T.; Shuker, S. B.; Meadows, R. P.; Steinman, D. H.; Carrera, G. M.; Marcotte, P. A.; Severin, J.; Walter, K.; Smith, H.; Gubbins, E.; Simmer, R.; Holzman, T. F.; Morgan, D. W.; Davidsen, S. K.; Summers, J. B.; Fesik, S. W. *J. Am. Chem. Soc.* **1997**, *119*, 5818–5827.
8. (a) Bodenhausen, G.; Ruben, D. J. *Chem. Phys. Lett.* **1980**, *69*, 185–189; (b) Kay, L. E.; Keifer, P.; Saarinen, T. *J. Am. Chem. Soc.* **1992**, *114*, 10663–10665.
9. Wold, S.; Albano, C.; Dunn, W. J.; Edlund, U.; Esbenson, K.; Geladi, P.; Hellberg, S.; Lindberg, W.; Sjöström, M. In *Chemometrics: Mathematics and Statistics in Chemistry*; Kowalski, B., Ed.; Reidel: Dortrecht, The Netherlands, 1984; pp 17–95.
10. Matter, H.; Schwab, W. *J. Med. Chem.* **1999**, *42*, 4506–4523.
11. Birkedal-Hansen, H.; Moore, W. G. I.; Bodden, M. K.; Windsor, L. J.; Birkedal-Hansen, B.; DeCarlo, A.; Engler, J. A. *Crit. Rev. Oral Biol. Med.* **1993**, *4*, 197–250.
12. Lohmander, L. S.; Hoerrner, L. A.; Lark, M. W. *Arthrit. Rheum.* **1993**, *36*, 181–185.
13. Murphy, G.; Hembry, R. M. *J. Rheumatol.* **1992**, *19*, 61–64.
14. Peress, N.; Perillo, E.; Zucker, S. *J. Neuropathol. Exp. Neurol.* **1995**, *54*, 16–22.
15. Schlechter, I.; Berger, A. *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157–162.
16. Kaplan, J. I.; Fraenkel, G. *NMR of Chemically Exchanging Systems*; Academic: NY, USA, 1980.
17. Lian, L. Y.; Roberts, G. C. K. In *NMR of Macromolecules*; Smith, D. W., Ed.; Oxford University Press: Oxford, UK, 1993; pp 153–182.
18. ^{15}N-labeled MMP-3 was produced in *E. coli* and the purified following Ref. 7. Samples were prepared from a 100 μM solution of Stromelysin in 20 mM BisTris buffer (pH 7.0) with 10 mM $CaCl_2$ and 100 mM acetohydroxamic acid (inhibitor) in water/D_2O (9:1). All spectra were taken on a BRUKER DRX 500 spectrometer operating at a basic frequency of 500.13 MHz for ^{1}H using a Cryo-ProbeTM. HSQC-spectra (Schleucher, J.; Schwendinger, M.; Sattler, M.; Schmidt, P.; Schedletzky, O.; Glaser, S. J.; Sorensen, O. W.; Griesinger, C. *J. Biomol. NMR* **1994**, *4*, 301–306) were recorded with a spectral width of 14 ppm (f2 = ^{1}H) and 34 ppm (f1 = ^{15}N) sampling 2048 complex data points along f2 and 128 real data points along f1. Protein samples were titrated with ligand solutions in DMSO (concentration > 50 mM) quantified prior to use by NMR-spectra taken with ammonium-formiate as an internal standard. Aliquots were added successively to the protein samples resulting in a final ligand concentration of 20, 40, 80, and 130 μM. For statistical analysis, spectra allowing for the detection of free and bound protein resonances were evaluated, typically at 40 μM ligand concentration.
19. IC_{50} values were determined as described before.[10,22]
20. Statistical analyses are based on GOLPE 4.5 (Baroni, M.; Costantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S. *Quant. Struct.-Act. Relat.* **1993**, *12*, 9–20). All columns (^{1}H, ^{15}N shift differences) with a low standard deviation (<0.01) were rejected. The remaining columns were normalized using autoscaling.
21. The PDB file 1caq (Pavlovsky, A. G.; Williams, M. G.; Ye, Q.-Z.; Ortwine, D. F.; Purchase, C. F.; White, A. D.; Dhanaraj, V.; Roth, B. D.; Johnson, L. L.; Hupe, D.; Humblet, C.; Blundell, T. L. *Protein Sci.* **1999**, *8*, 1455–1462) was used for modeling. Energy calculations were based on the MMFF94s force field (Halgren, T. J. *Comp. Chem.* **1999**, *20*, 720–729). For docking, ligands were placed in the site and optimized treating all side chains within 4 Å flexible. Automated docking was carried out using FlexiDock (Tripos) and QXP (McMartin, C.; Bohacek, R. S. *J. Comput. Aided Mol. Des.* **1997**, *11*, 333–344).
22. Matter, H.; Schwab, W.; Barbier, D.; Billen, G.; Haase, B.; Neises, B.; Schudok, M.; Thorwart, W.; Schreuder, H.; Brachvogel, V.; Lönze, P.; Weithmann, K.-U. *J. Med. Chem.* **1999**, *42*, 1908–1920.