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.

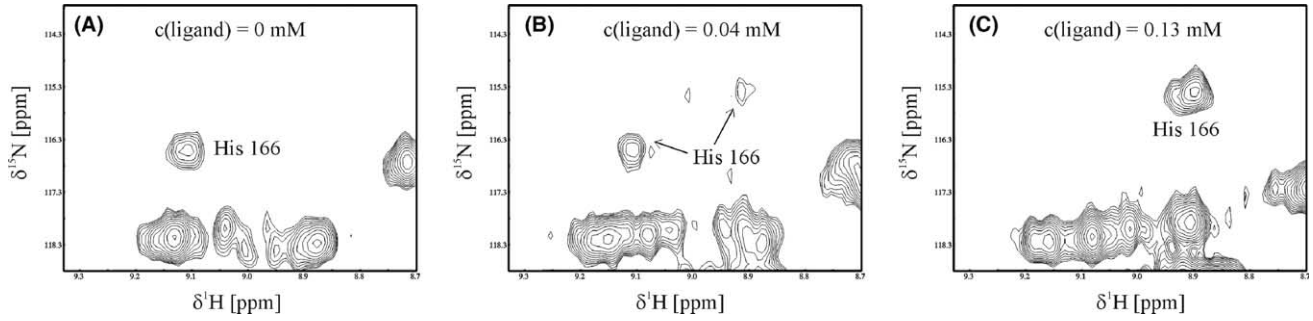

Figure 1. Selected regions from 500 MHz $^{1}H/15N$ HSQC-spectra for uncomplexed MMP-3 (A) and its complex with inhibitor **62** (IC_{50} 500 nM) in different concentrations (B, C). As high-affinity results in slow exchange, two sets of signals are observed for uncomplexed and complexed MMP-3, as exemplified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