List of References for COMBINE Analysis

21.01.2009

Arakawa, M., K. Hasegawa, et al. (2008). "Tailored scoring function of Trypsin–benzamidine complex using COMBINE descriptors and support vector regression." Chemometrics and Intelligent Laboratory Systems **92**(2): 145-151.

> Structure-based drug design (SBDD) is a computational technique for designing new drug candidates based on physico-chemical interactions between a protein and a ligand molecule. The most important thing for SBDD is accurate estimation of binding affinity of the ligand molecule against the target protein. Scoring function, which is basically a mathematical equation that approximates the thermodynamics of binding, has to be defined in advance. In this paper, we propose a novel method for building a tailored scoring function using comparative molecular binding energy (COMBINE) descriptors and support vector regression (SVR). COMBINE descriptors are energy terms between the ligand molecule and each amino acid residue of the target protein. SVR is a promising nonlinear regression method based on the theory of support vector machine (SVM). In these types of regression methodology, variable selection is one of the most important issues to construct a robust and predictive quantitative structure–activity relationship (QSAR) model. We adopted a variable selection method based on sensitivity analysis of each variable. The usefulness of the proposed method has been validated by applying to real QSAR data set, benzamidine derivatives as Trypsin inhibitors. The final SVR model could successfully identify important amino acid residues for explaining inhibitory activities.

Cuevas, C., M. Pastor, et al. (2001). "Comparative binding energy (COMBINE) analysis of human neutrophil elastase inhibition by pyridone-containing trifluoromethylketones." Comb Chem High Throughput Screen **4**(8): 627-42.

> The complexes of human neutrophil elastase with a series of 40 N3-substituted trifluoromethylketone-based pyridone inhibitors have been modelled. The series spans three orders of magnitude in inhibition constants despite the fact that it was originally developed in an attempt to improve the oral activity of a lead compound. Ligand-receptor interaction energies calculated using molecular mechanics did not correlate well with the experimental activities. A good correlation with activity was found, however, when a COMBINE analysis of the same data was carried out, which allowed a quantitative interpretation of the modelled complexes. The essence of this method is to partition the ligand-receptor interaction energies into individual residue-based van der Waals and electrostatic contributions, and to subject the resulting energy matrix to partial least squares analysis. Incorporation of two additional descriptors representing the electrostatic energy contributions to the partial desolvation of both the receptor and the ligands improved the QSAR model, as did the replacement of the distance-dependent electrostatic contributions with solvent-screened electrostatic interactions calculated by numerically solving the Poisson-Boltzmann equation. The model was validated both internally (cross-validation) and externally, using a set of twelve 6-phenyl-pyridopyrimidine analogs. The analysis reveals the subtle interplay of binding forces which occurs within the enzyme active site and provides objective information that can be interpreted in the light of the receptor structure. This information, gained from a series of real compounds, can be easily translated into 3D real or virtual database queries in the search for more active derivatives.

Damborsky, J., J. Kmunicek, et al. (2004). Rational re-design of haloalkane dehalogenases guided by comparative binding energy analysis. Enzyme Functionality: Design, Engineering and Screening. A. Svendsen and M. Dekker. New York, ISBN:0-8247-4709-7**:** 79-96.

Gollapudy, R., S. Ajmani, et al. (2004). "Modeling and interactions of Aspergillus fumigatus lanosterol 14-alpha demethylase 'A' with azole antifungals." Bioorg Med Chem **12**(11): 2937-50.

Recent identification of the sterol 14-alpha demethylase genes (CYP51 A and B) from Aspergillus fumigatus and other species by Mellado et al. (J. Clin. Microbiol. 2001, 39(7), 2431-2438), has opened up possibilities of investigating the interactions of azole antifungals with the enzyme(s) from fungi. This study describes for the first time, a model of the three-dimensional structure of A. fumigatus 14-alpha demethylase (AF-CYP51A), using the crystal structure of Mycobacterium tuberculosis 14-alpha demethylase (PDB code:1EA1) as a template. The paper also describes the various interactions between azole antifungals and the target from A. fumigatus (AF-CYP51A). Quantitative evaluation of these interactions is done using COMBINE analysis to understand contributions of active site residues to ligand activity. It also provides explanation for the activity/inactivity of different ligands for AF-CYP51A.

Guo, J., M. M. Hurley, et al. (2004). "A Docking Score Function for Estimating Ligand-Protein Interactions: Application to Acetylcholinesterase Inhibition." J. Med. Chem. **47**(22): 5492-5500.

Abstract: Acetylcholinesterase (AChE) inhibition is an important research topic because of its wide range of associated health implications. A receptor-specific scoring function was developed herein for predicting binding affinities for human AChE (huAChE) inhibitors. This method entails a statistically trained weighted sum of electrostatic and van der Waals (VDW) interactions between ligands and the receptor residues. Within the 53 ligand training set, a strong correlation was found (R2 = 0.89) between computed and experimental inhibition constants. Leave-one-out cross-validation indicated high predictive power (Q2 = 0.72), and analysis of a separate 16-compound test set also produced very good correlation with experiment (R2 = 0.69). Scoring function analysis has permitted identification and characterization of important ligand-receptor interactions, producing a list of those residues making the most important electrostatic and VDW contributions within the main active site, gorge area, acyl binding pocket, and periferal site. These analyses are consistent with X-ray crystallographic and site-directed mutagenesis studies.

Hasegawa, K., T. Kimura, et al. (1999). "GA Strategy for Variable Selection in QSAR Studies: Enhancement of Comparative Molecular Binding Energy Analysis by GA-Based PLS Method." Quantitative Structure-Activity Relationships **18**(3): 262-272.

Comparative molecular binding energy (COMBINE) is a novel approach for estimation of binding affinity in structure-based drug design (SBDD). COMBINE involves an extensive partitioning of binding interaction energy and multivariate regression analysis to derive a model. In COMBINE, partial least squares (PLS) is especially used as a statistical method. Although PLS is robust and stable, it has been shown that its predictive performance drops with the increase of number of variables. Also, from a practical point of view, model becomes complicated and its interpretation is difficult if we use many variables. Therefore, it is expected that PLS coupled with variable selection can produce a more predictive and interpretable model in COMBINE. The purpose of this paper is to examine whether genetic algorithm-based PLS (GAPLS) developed by our group for variable selection can enhance prediction

and interpretation of the COMBINE model. The structure-activity data of human immuno-deficiency virus type I (HIV-1) protease inhibitors were used as a test example. By applying GAPLS to this data set, several improved PLS models with a high cross-validated r2 value and low number of variables were obtained. In order to select a best model from them, external validation was performed for each model. The finally selected model was further examined by comparing with the 3D structure of HIV-1 protease in computer graphics and its agreement was confirmed.

Kim, H. J., C. H. Chae, et al. (2004). "Computational studies of COX-2 inhibitors: 3D-QSAR and docking." Bioorg Med Chem **12**(7): 1629-41.

The 3D-QSAR (three-dimensional quantitative structure-activity relationships) studies for 88 selective COX-2 (cyclooxygenase-2) inhibitors belonging to three chemical classes (triaryl rings, diaryl cycloalkanopyrazoles, and diphenyl hydrazides) were conducted using comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA). Partial least squares analysis produced statistically significant models with q(2) values of 0.84 and 0.79 for CoMFA and CoMSIA, respectively. The binding energies calculated from flexible docking were correlated with inhibitory activities by the least-squares fit method. The three chemical classes of inhibitors showed reasonable internal predictability (r(2)=0.51, 0.49, and 0.54), but the sulfonyl-containing inhibitors demonstrated distinctively low binding energy compared to the others. The electrostatic interaction energy between the Arg513 of the COX-2 active site and sulfonyl group of the triaryl rings seemed to have the responsibility for difference in binding energy. Comparative binding energy (COMBINE) analyses gave q(2) values of 0.64, 0.63, and 0.50 for triaryl rings, diaryl cycloalkanopyrazoles, and diphenyl hydrazides, respectively. In this COMBINE model, some protein residues were highlighted as particularly important for inhibitory activity. The combination of ligand-based and structure-based models provided an improved understanding in the interaction between the three chemical classes and the COX-2.

Kmunicek, J., M. Bohac, et al. (2003). "Comparative binding energy analysis of haloalkane dehalogenase substrates: modelling of enzyme-substrate complexes by molecular docking and quantum mechanical calculations." J Comput Aided Mol Des **17**(5-6): 299-311.

We evaluate the applicability of automated molecular docking techniques and quantum mechanical calculations to the construction of a set of structures of enzyme-substrate complexes for use in Comparative binding energy (COMBINE) analysis to obtain 3D structure-activity relationships. The data set studied consists of the complexes of eighteen substrates docked within the active site of haloalkane dehalogenase (DhlA) from Xanthobacter autotrophicus GJ10. The results of the COMBINE analysis are compared with previously reported data obtained for the same dataset from modelled complexes that were based on an experimentally determined structure of the DhlA-dichloroethane complex. The quality of fit and the internal predictive power of the two COMBINE models are comparable, but better external predictions are obtained with the new approach. Both models show a similar composition of the principal components. Small differences in the relative contributions that are assigned to important residues for explaining binding affinity differences can be directly linked to structural differences in the modelled enzyme-substrate complexes: (i) rotation of all substrates in the active site about their longitudinal axis, (ii) repositioning of the ring of epihalohydrines and the halogen substituents of 1,2-dihalopropanes, and (iii) altered conformation of the long-chain molecules (halobutanes and halohexanes). For external validation, both a novel

substrate not included in the training series and two different mutant proteins were used. The results obtained can be useful in the future to guide the rational engineering of substrate specificity in DhlA and other related enzymes.

Kmunicek, J., K. Hynkova, et al. (2005). "Quantitative analysis of substrate specificity of haloalkane dehalogenase LinB from Sphingomonas paucimobilis UT26." Biochemistry **44**(9): 3390-401.

Haloalkane dehalogenases are microbial enzymes that cleave a carbon-halogen bond in halogenated compounds. The haloalkane dehalogenase LinB, isolated from Sphingomonas paucimobilis UT26, is a broad-specificity enzyme. Fifty-five halogenated aliphatic and cyclic hydrocarbons were tested for dehalogenation with the LinB enzyme. The compounds for testing were systematically selected using a statistical experimental design. Steady-state kinetic constants K(m) and k(cat) were determined for 25 substrates that showed detectable cleavage by the enzyme and low abiotic hydrolysis. Classical quantitative structure-activity relationships (QSARs) were used to correlate the kinetic constants with molecular descriptors and resulted in a model that explained 94% of the experimental data variability. The binding affinity of the tested substrates for this haloalkane dehalogenase correlated with hydrophobicity, molecular surface, dipole moment, and volume:surface ratio. Binding of the substrate molecules in the active site pocket of LinB depends nonlinearly on the size of the molecules. Binding affinity increases with increasing substrate size up to a chain length of six carbon atoms and then decreases. Comparative binding energy (COMBINE) analysis was then used to identify amino acid residues in LinB that modulate its substrate specificity. A model with three statistically significant principal components explained 95% of the experimental data variability. van der Waals interactions between substrate molecules and the enzyme dominated the COMBINE model, in agreement with the importance of substrate size in the classical QSAR model. Only a limited number of protein residues (6-8%) contribute significantly to the explanation of variability in binding affinities. The amino acid residues important for explaining variability in binding affinities are as follows: (i) first-shell residues Asn38, Asp108, Trp109, Glu132, Ile134, Phe143, Phe151, Phe169, Val173, Trp207, Pro208, Ile211, Leu248, and His272, (ii) tunnel residues Pro144, Asp147, Leu177, and Ala247, and (iii) second-shell residues Pro39 and Phe273. The tunnel and the second-shell residues represent the best targets for modulating specificity since their replacement does not lead to loss of functionality by disruption of the active site architecture. The mechanism of molecular adaptation toward a different specificity is discussed on the basis of quantitative comparison of models derived for two protein family members.

Kmunicek, J., S. Luengo, et al. (2001). "Comparative binding energy (COMBINE) analysis of the substrate specificity of haloalkane dehalogenase from xanthobacter autotrophicus GJ10." Biochemistry **40**: 8905-8917.

Lozano, J. J., M. Pastor, et al. (2000). "3D-QSAR methods on the basis of ligand-receptor complexes. Application of COMBINE and GRID/GOLPE methodologies to a series of CYP1A2 ligands." J Comput Aided Mol Des **14**(4): 341-53.

Many heterocyclic amines (HCA) present in cooked food exert a genotoxic activity when they are metabolised (N-oxidated) by the human cytochrome P450 1A2 (CYP1A2h). In order to rationalize the observed differences in activity of this enzyme on a series of 12 HCA, 3D-QSAR methods were applied on the basis of models of HCA-CYP1A2h complexes. The CYP1A2h enzyme model has been previously

reported and was built by homology modeling based on cytochrome P450 BM3. The complexes were automatically generated applying the AUTODOCK software and refined using AMBER. A COMBINE analysis on the complexes identified the most important enzyme-ligand interactions that account for the differences in activity within the series. A GRID/GOLPE analysis was then performed on just the ligands, in the conformations and orientations found in the modeled complexes. The results from both methods were concordant and confirmed the advantages of incorporating structural information from series of ligand-receptor complexes into 3D-QSAR methodologies.

Lushington, G. H., J. X. Guo, et al. (2007). "Whither combine? New opportunities for receptor-based QSAR." Curr Med Chem **14**(17): 1863-77.

Receptor based QSAR methods represent a computational marriage of structure activity relationship analysis and receptor structure based design that is providing valuable pharmacological insight to a wide range of therapeutic targets. One implementation, called Comparative Binding Energy (COMBINE) analysis, is particularly powerful by virtue of its explicit consideration of interatomic interactions between the ligand and receptor as the QSAR variable space. This review outlines the methodological basis for the COMBINE model, contrasts it relative to other 3D QSAR techniques, and discusses sample applications that illustrate recent key innovations. One major development discussed is the rigorous integration of multiple receptors into unified COMBINE models for probing bioactivity trends as a function of amino acid variation across a series of homologous protein receptors, and as a function of conformational variation within one single protein. Other important examples include a recent extension of the method to account for covalent effects arising from ligand binding, as well as successful application of a COMBINE model to high throughput virtual screening. This review concludes with discussions about possible future methodological refinements and their applications, including potential extensions to four-dimensional QSAR, and a potential role of quantum chemistry in addressing covalent bonding effects and parametric adaptivity.

Martin-Santamaria, S., J. Munoz-Muriedas, et al. (2004). "Modulation of binding strength in several classes of active site inhibitors of acetylcholinesterase studied by comparative binding energy analysis." J Med Chem **47**(18): 4471-82.

The comparative binding energy (COMBINE) methodology has been used to identify the key residues that modulate the inhibitory potencies of three structurally different classes of acetylcholinesterase inhibitors (tacrines, huprines, and dihydroquinazolines) targeting the catalytic active site of this enzyme. The extended set of energy descriptors and the partial least-squares methodology used by COMBINE analysis on a unique training set containing all the compounds yielded an interpretable model that was able to fit and predict the activities of the whole series of inhibitors reasonably well ($r2 = 0.91$ and $q2 = 0.76$, 4 principal components). A more robust model ($q2 = 0.81$ and SDEP = 0.25, 3 principal components) was obtained when the same chemometric analysis was applied to the huprines set alone, but the method was unable to provide predictive models for the other two families when they were treated separately from the rest. This finding appears to indicate that the enrichment in chemical information brought about by the inclusion of different classes of compounds into a single training set can be beneficial when an internally consistent set of pharmacological data can be derived. The COMBINE model was externally validated when it was shown to predict the activity of an additional set of compounds that were not employed in model construction. Remarkably, the differences in inhibitory

potency within the whole series were found to be finely tuned by the electrostatic contribution to the desolvation of the binding site and a network of secondary interactions established between the inhibitor and several protein residues that are distinct from those directly involved in the anchoring of the ligand. This information can now be used to advantage in the design of more potent inhibitors.

Murcia, M., A. Morreale, et al. (2006). "COMBINE analysis considering multiple receptors: a step towards structure-activity models for protein families." J. Med. Chem. **49**: 6241-6253.

Murcia, M. and A. R. Ortiz (2004). "Virtual screening with flexible docking and COMBINE-based models. Application to a series of factor Xa inhibitors." J Med Chem **47**(4): 805-20.

A two-step, fully automatic virtual screening procedure consisting of flexible docking followed by activity prediction by COMparative BINding Energy (COMBINE) analysis is presented. This novel approach has been successfully applied, as an example with medicinal chemistry interest, to a recently reported series of 133 factor Xa (fXa)(1) inhibitors whose activities encompass 4 orders of magnitude. The docking algorithm is linked to the COMBINE analysis program and used to derive independent regression models of the 133 inhibitors docked within three different fXa structures (PDB entries 1fjs, 1f0r, and 1xka), so as to explore the effect of receptor conformation on the overall results. Reliable docking conformations and predictive regression models requiring eight latent variables could be derived for two of the fXa structures, with the best model achieving a Q(2) of 0.63 and a standard deviation of errors of prediction (SDEP) of 0.51 (leave-one-out). The two-step procedure was then employed to screen a designed virtual library of 112 ligands, containing both active and inactive compounds. While docking energies alone could show a good performance for selecting hits, including structurally diverse ones, inclusion of COMBINE analysis regression models provided improved rankings for the identification of structurally related molecules in external sets. In our best case, a recognition rate of approximately 80% of known binders at approximately 15% false positives rate was achieved, corresponding to an enrichment factor of approximately 450% over random.

Nakamura, S., I. Nakanishi, et al. (2006). "Binding affinity prediction of non-peptide inhibitors of HIV-1 protease using COMBINE model introduced from peptide inhibitors." Bioorg Med Chem Lett **16**(24): 6334-6337.

Comparative binding energy (COMBINE) analysis method is one of the QSAR techniques for the prediction of biological activities of inhibitors based on interaction energies between ligands and proteins decomposed into each amino acid residue. We supposed that the predictive ability of the COMBINE method does not depend essentially on the molecular frameworks of ligands. To verify this idea, we performed the COMBINE analysis of non-peptide inhibitors of HIV-1 protease (HIVp), where the prediction model was constructed using inhibitors with a peptide scaffold as a training set. The predictive performance of the AMBER and CHARMm force fields was very high and at the same level (q(2)=0.75, 0.67, SDEP(cv)=0.76, 0.89, and SDEP(ex)=0.92, 0.66, respectively). The high predictive ability of the COMBINE method for the distinct scaffold compounds is due to the informative description of the interaction energies for compounds that are located at the binding site. This result suggests that COMBINE analysis may be applied not only to the lead optimization stage but also to the lead evolution stages.

Ortiz, A. R., M. Pastor, et al. (1997). "Reliability of comparative molecular field analysis models: effects of data scaling and variable selection using a set of human synovial fluid phospholipase A2 inhibitors." J Med Chem. **40**: 1136-1148.

Ortiz, A. R., M. T. Pisabarro, et al. (1995). "Prediction of drug binding affinities by comparative binding energy analysis." J. Med. Chem. **38**: 2681-2691.

Pastor, M., F. Gago, et al. (2000). Comparative binding energy (COMBINE) analysis on a series of glycogen phosphorylase inhibitors: comparison with GRID/GOLPE methods. Molecular Modeling and Prediction of Bioactivity. K. Gundertofter and F. S. Jorgensen. New York, Kluwer**:** 329-330.

Pastor, M., C. Perez, et al. (1997). "Simulation of alternative binding modes in a structure-based QSAR study of HIV-1 protease inhibitors." J Mol Graph Model **15**(6): 364-71, 389.

We have used a published set of inhibitors of HIV-1 protease to build a COMBINE-type structure-based QSAR model with good predictive ability (r2 = 0.90, q2 = 0.69). Since the compounds in the training series exhibit most of their structural variability on one-half of the pseudosymmetrical binding cavity and only one binding orientation was explored for each molecule, the model describes mainly the effect of the structural changes on interactions involving only one-half of the binding cavity (pockets S1' and S2'). Thus, the model cannot be expected to give accurate predictions for new compounds exhibiting structural variation in both halves. The model does in fact show a tendency to underpredict slightly the biological activity of the molecules in the external test set. In an attempt to improve the quality of the model, both possible orientations of the ligands are now considered so that structural variation takes place in all binding pockets. One possibility would have been to build an additional set of complexes with the inhibitors docked in a reversed orientation. The alternative we have explored, however, consists of manipulating the data matrix describing the interaction energies so that each row is duplicated and the order of the variables in the duplicated rows is swapped between subunits. This simple approach has produced a new model that is similar in quality to the original model (r2 = 0.89, q2 = 0.64) but lacks the tendency to underpredict the activity of the compounds in the external set. Moreover, since equivalent residues are assigned equivalent weights, the model is insensitive to ligand orientation and is easier to interpret.

Perez, C., M. Pastor, et al. (1998). "Comparative binding energy analysis of HIV-1 protease inhibitors: incorporation of solvent effects and validation as a powerful tool in receptor-based drug design." J. Med. Chem. **41**: 836-852.

Peters, M. B. and K. M. Merz (2006). "Semiempirical Comparative Binding Energy Analysis (SE-COMBINE) of a Series of Trypsin Inhibitors." J. Chem. Theory Comput. **2**(2): 383-399.

Abstract: A scheme to decompose the intermolecular interaction energy of a series of complexes at the semiempirical (SE) level has been developed and validated. The comparative binding energy analysis (COMBINE) (Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C. J. Med. Chem. 1995, 38, 2681-2691) and the semiempirical quantum mechanical method pairwise energy decomposition (PWD) (Raha, K.; van der Vaart, A. J.; Riley, K. E.; Peters, M. B.; Westerhoff, L. M. Kim, H.; Merz, K. M., Jr. J. Am. Chem. Soc. 2005, 127, 6583-6594) were coupled together to form SE-COMBINE. This approach calculates the residue pairwise electrostatic interaction energies, and QSAR models were built with the energies as descriptors using partial least squares (PLS). The application of SE-COMBINE was used as an investigation of

the intermolecular interactions between 88 benzamidine inhibitors and trypsin and to test the ability of this new method to predict binding free energies. The predictive capability of SE-COMBINE is shown to be comparable to those of other QSAR methods, and using graphical intermolecular interaction maps (IMMs) enhances the interpretability of receptor-based QSARs.

Rodriguez-Barrios, F. and F. Gago (2004). "Chemometrical identification of mutations in HIV-1 reverse transcriptase conferring resistance or enhanced sensitivity to arylsulfonylbenzonitriles." J Am Chem Soc **126**(9): 2718-9.

Schleinkofer, K., U. Wiedemann, et al. (2004). "Comparative structural and energetic analysis of WW domain-peptide interactions." J Mol Biol **344**(3): 865-81.

WW domains are small globular protein interaction modules found in a wide spectrum of proteins. They recognize their target proteins by binding specifically to short linear peptide motifs that are often proline-rich. To infer the determinants of the ligand binding propensities of WW domains, we analyzed 42 WW domains. We built models of the 3D structures of the WW domains and their peptide complexes by comparative modeling supplemented with experimental data from peptide library screens. The models provide new insights into the orientation and position of the peptide in structures of WW domain-peptide complexes that have not yet been determined experimentally. From a protein interaction property similarity analysis (PIPSA) of the WW domain structures, we show that electrostatic potential is a distinguishing feature of WW domains and we propose a structure-based classification of WW domains that expands the existent ligand-based classification scheme. Application of the comparative molecular field analysis (CoMFA), GRID/GOLPE and comparative binding energy (COMBINE) analysis methods permitted the derivation of quantitative structure-activity relationships (QSARs) that aid in identifying the specificity-determining residues within WW domains and their ligand-recognition motifs. Using these QSARs, a new group-specific sequence feature of WW domains that target arginine-containing peptides was identified. Finally, the QSAR models were applied to the design of a peptide to bind with greater affinity than the known binding peptide sequences of the yRSP5-1 WW domain. The prediction was verified experimentally, providing validation of the QSAR models and demonstrating the possibility of rationally improving peptide affinity for WW domains. The QSAR models may also be applied to the prediction of the specificity of WW domains with uncharacterized ligand-binding properties.

Tomic, S., B. Bertosa, et al. (2007). "COMBINE analysis of the specificity of binding of Ras proteins to their effectors." Proteins: Structure, Function, and Bioinformatics **67**(2): 435-447.

The small guanosine triphosphate (GTP)-binding proteins of the Ras family are involved in many cellular pathways leading to cell growth, differentiation, and apoptosis. Understanding the interaction of Ras with other proteins is of importance not only for studying signalling mechanisms but also, because of their medical relevance as targets, for anticancer therapy. To study their selectivity and specificity, which are essential to their signal transfer function, we performed COMparative BINding Energy (COMBINE) analysis for 122 different wild-type and mutant complexes between the Ras proteins, Ras and Rap, and their effectors, Raf and RalGDS. The COMBINE models highlighted the amino acid residues responsible for subtle differences in binding of the same effector to the two different Ras proteins, as well as more significant differences in the binding of the two different effectors (RalGDS and Raf) to Ras. The study revealed that E37, D38, and D57 in Ras are

nonspecific hot spots at its effector interface, important for stabilization of both the RalGDS-Ras and Raf-Ras complexes. The electrostatic interaction between a GTP analogue and the effector, either Raf or RalGDS, also stabilizes these complexes. The Raf-Ras complexes are specifically stabilized by S39, Y40, and D54, and RalGDS-Ras complexes by E31 and D33. Binding of a small molecule in the vicinity of one of these groups of amino acid residues could increase discrimination between the Raf-Ras and RalGDS-Ras complexes. Despite the different size of the RalGDS-Ras and Raf-Ras complexes, we succeeded in building COMBINE models for one type of complex that were also predictive for the other type of protein complex. Further, using system-specific models trained with only five complexes selected according to the results of principal component analysis, we were able to predict binding affinities for the other mutants of the particular Ras-effector complex. As the COMBINE analysis method is able to explicitly reveal the amino acid residues that have most influence on binding affinity, it is a valuable aid for protein design. Proteins 2007. © 2007 Wiley-Liss, Inc.

Tomic, S. and B. Kojic-Prodic (2002). "A quantitative model for predicting enzyme enantioselectivity: Application to burkholderia cepacia lipase and 3-(aryloxy)-1,2-propanediol derivatives." J. Mol. Graph. Model. **21**(3): 241-252.

Tomic, S., L. Nilsson, et al. (2000). "Nuclear receptor-DNA binding specificity: A COMBINE and Free-Wilson QSAR analysis." J Med Chem **43**(9): 1780-92.

Specific binding of transcription factors to DNA is crucial for gene regulation. We derived models for the binding specificity of transcription factors of the nuclear receptor family to DNA using two QSAR methods: a Free-Wilson-like method and COMparative BINding Energy (COMBINE) analysis. The analysis is based on experimental data for the interaction of 20 mutant glucocorticoid receptor DNA-binding domains with 16 different response elements in a total of 320 complexes (Zilliacus, J.; Wright, A. P.; Carlstedt-Duke, J.; Nilsson, L.; Gustafsson, J. A. Proteins 1995, 21, 57-67). The predictive abilities of the models obtained by the two methods are similar. The COMBINE analysis indicates that the most important properties for determining binding specificity for this dataset are the changes upon binding of the solvation free energies of the bases that are mutated in the dataset and the electrostatic interactions of the mutated nucleotides with certain charged amino acids. Further important descriptors are the changes of solvation free energy and surface area of the side chain of the mutated residue. It is clear, however, that there are additional features important for the specificity of binding that are not included in the models, such as differences in interfacial hydration of the complexes.

Tomic, S. and R. C. Wade (2001). "COMBINE analysis of nuclear receptor-DNA binding specificity: Comparison of two datasets." Croat. Chem. Acta **74**: 295-314.

Wade, R. C. (2001). Derivation of QSARs using 3D structural models of protien-ligand complexes by COMBINE analysis. Rational Approaches to Drug Design: 13th European Symposium on Quantitative Structure-Activity Relationships. H.-D. Holtje and W. Sippl. Barcelona, Prous Science S. A.**:** 23-28.

With the accelerating pace of protein structure determination and the advent of structural genomics programmes, it will be more and more common to know the 3D protein structures of drug design targets. This will increase the need to be able to derive QSARs that are based on information in protein 3D structures. With the goal of deriving QSARs in an objective and quantitative way from 3D models of protein-

ligand complexes, we developed the COMparative BINding Energy (COMBINE) Analysis method (1,2). In this talk, I will describe recent developments and applications of COMBINE analysis.

To carry out a COMBINE analysis, a series of protein-ligand complexes of known activity is modelled and energy-minimized. The molecular mechanics interaction energy between protein and ligand is computed for each complex and partitioned into terms according to the physical nature of the interaction (e.g. Coulombic and Lennard-Jones) and the location in each molecule (e.g. on a per residue basis). Then a PLS analysis is performed to derive a QSAR model identifying the most important energy terms for the activity. This QSAR can be used to make predictions for new molecules or for protein mutants whose complexes are modelled in the same way. In addition to intermolecular interaction energies, other energy terms can be treated in the derivation of the COMBINE QSAR, including changes upon binding in terms describing intramolecular energy, conformational entropy and desolvation energy. COMBINE analysis has been applied to derive QSAR models for enzyme inhibition (2-6), enzyme-substrate binding (7-8) and protein receptor-DNA binding (9).

Wade, R. C., S. Henrich, et al. (2004). "Using 3D protein structures to derive 3D-QSARs." Drug Discovery Today: Technologies **1**(3): 241-246.

The three-dimensional structures of proteins are being solved apace, yet this information is often underused in quantitative structure-activity relationship (QSAR) studies. Here, we describe and compare methods for exploiting protein structures to derive 3D-QSARs. These methods can facilitate molecular design and lead optimization and should increasingly become a standard component of the drug designer's repertoire.

Wade, R. C., A. R. Oritz, et al. (1998). "Comparative binding energy analysis." Perspectives in Drug Discovery and Design **9**: 19-34.

Wang, T., S. Tomic, et al. (2004). "How optimal are the binding energetics of barnase and barstar?" Biophys J **87**(3): 1618-30.

The extracellular ribonuclease barnase and its intracellular inhibitor barstar bind fast and with high affinity. Although extensive experimental and theoretical studies have been carried out on this system, it is unclear what the relative importance of different contributions to the high affinity is and whether binding can be improved through point mutations. In this work, we first applied Poisson-Boltzmann electrostatic calculations to 65 barnase-barstar complexes with mutations in both barnase and barstar. The continuum electrostatic calculations with a van der Waals surface dielectric boundary definition result in the electrostatic interaction free energy providing the dominant contribution favoring barnase-barstar binding. The results show that the computed electrostatic binding free energy can be improved through mutations at W44/barstar and E73/barnase. Furthermore, the determinants of binding affinity were quantified by applying COMparative BINding Energy (COMBINE) analysis to derive quantitative structure-activity relationships (QSARs) for the 65 complexes. The COMBINE QSAR model highlights approximately 20 interfacial residue pairs as responsible for most of the differences in binding affinity between the mutant complexes, mainly due to electrostatic interactions. Based on the COMBINE model, together with Brownian dynamics simulations to compute diffusional association rate constants, several mutants were designed to have higher binding affinities than the wild-type proteins.

Wang, T. and R. C. Wade (2001). COMBINE 3D-QSAR analysis of influenza neuraminidase inhibitors. Rational Approaches to Drug Design: 13th European Symposium on Quantitative Structure-Activity Relationships. H.-D. Holtje and W. Sippl. Barcelona, Prous Science S. A.**:** 78-82.

Neuraminidase (NA) is a surface glycoprotein of influenza viruses which cleaves terminal sialic acids from glycoproteins and glycolipids and is critical for viral replication. The active site of neuraminidase[1,2] is conserved in all type A and B influenza viruses, making it an excellent target for anti-influenza drug design. Indeed, neuraminidase inhibitors have recently become available in the clinic for treatment of influenza (http://www.glaxowellcome.co.uk/fighting). We describe the use of structures of protein-inhibitor complexes to derive quantitative structure-activity relationships (QSARs) which should aid understanding of the mechanism of inhibition and the discovery of new inhibitors. Crystal structures of 30 NA-inhibitor complexes as well as 12 complexes with inhibitors docked by the AUTODOCK program[3] were used to build a 3D-QSAR model by COMparative BINding Energy (COMBINE) analysis[4]. The proteins included subtype N2, subtype N9 and the N9 mutant of type A NA. The inhibitors [5-6] included sialic acid and benzoic acid analogues. A predictive QSAR model was obtained and protein residues and bound water molecules important for inhibitor activity were highlighted in the QSAR model. In addition, based on the COMBINE analysis, a 4-point pharmacophore was proposed and was used to search for matching compounds in 3D structural databases by the 3DFS program [7]. Some hits with novel structures were obtained.

Wang, T. and R. C. Wade (2001). "Comparative binding energy (COMBINE) analysis of influenza neuraminidase-inhibitor complexes." J Med Chem **44**: 961-671.

Wang, T. and R. C. Wade (2002). "Comparative binding energy (COMBINE) analysis of OppA-peptide complexes to relate structure to binding thermodynamics." J. Med. Chem. **45**: 4828-4837.