Anchor−GRIND: Filling the Gap between Standard 3D QSAR and the GRid-INdependent Descriptors

Fabien Fontaine, Manuel Pastor,* Ismael Zamora, and Ferran Sanz

Research Unit on Biomedical Informatics (GRIB), IMIM / Universitat Pompeu Fabra, C / Dr. Aiguader, 80, E-08003 Barcelona, Spain

Received November 4, 2004

The aim of this work is to present the anchor−GRIND methodology. Anchor-GRIND efficiently combines a priori chemical and biological knowledge about the studied compounds with alignment-independent molecular descriptors derived from molecular interaction fields. Such descriptors are particularly useful for series of ligands sharing a common scaffold but with very diverse substituents. The method uses a specific position of the molecular structure (the "anchor point") to compare the spatial distribution of the molecular interaction fields of the substituents. The descriptors produced are more detailed and specific than the original GRIND while still avoiding the bias introduced by the alignment. Three data sets have been studied to demonstrate the usefulness of the anchor−GRIND methodology for 3D-QSAR modeling. The two first data sets respectively include congeneric series of the hepatitis C virus NS3 protease and of the acetylcholinesterase inhibitors. The third data set discriminates between factor Xa inhibitors of high and low affinity. In all the series presented, the models obtained with the anchor−GRIND are statistically sound and easy to interpret.

Introduction

Nowadays, 3D-QSAR methods such as CoMFA[1] are standard tools forming part of the basic arsenal commonly used in medicinal chemistry projects. However, they are criticized for the large amount of expert time required to obtain a structure−activity model. With the increased use of new synthetic technologies, such as solid-phase synthesis, the increasing rate of available compounds for biological testing requires faster techniques for analyzing the compounds and rationalizing the structure−activity relationships. In this context, alignment-independent descriptors constitute an appealing alternative to more time-consuming 3D-QSAR methodologies, since they can be run with minimal human supervision. In this study, we present a new class of alignment-free molecular descriptors, namely, anchor−GRIND. The *gr*id-*in*dependent *d*escriptors (GRIND) encode the spatial distribution of the molecular interaction fields (MIF) of the studied compounds.[2] In the anchor−GRIND method, to compare the MIF distribution of different compounds, the user defines a single common position in the structure of all the compounds in the series, so-called the "anchor point". This anchor point does not provide enough geometrical constrains to align the compounds studied; however, it is used by the method as a common reference point, making it is possible to describe the geometry of the MIF regions in a more precise way than GRIND does.

The anchor point is particularly easy to assign in data sets having some chemical substituents well-known as being crucial for the activity. For example, in a series of statin-based aspartyl protease inhibitors, the central hydroxyl group is a good anchor point, whereas for the aminergic GPCR antagonists, the anchor point would be the charged amine nitrogen. By providing this common point as reference for comparison, the specificity and the simplicity of the descriptors is improved over other general alignment-free methods such as GRIND[2] or VolSurf,[3] and the method places itself midway between alignment-dependent and alignment-independent methods. The applicability of the method is only hampered by the need to work on series sharing a common scaffold, which is nevertheless the case in a lot of medicinal chemistry data sets.

In fact, the idea of defining an anchor point from the topology of the compounds is not new. Mason et al.[4] defined the centroid of "privileged substructures" as a special pharmacophoric point for their four-point pharmacophore fingerprints. The result is a better pharmacophoric description of the compounds around a common specific substructure. The first use of an anchor point to describe differences between MIF was done for the cytochrome P450 2C9 binding site.[5] There, Zamora et al. described a receptor binding site by first defining the oxygen atom attached to the heme iron as the anchor point. Then, the binding site was described in terms of its relative geometry with respect to this anchor point, and such description was compared with the possible geometries of ligands, inserted in different orientations, to obtain the position of the ligand more likely to be oxidized by the cytochrome.

In this work, we have selected three cases where the anchor point approach is possible. Two data sets, a series of hepatitis C virus (HCV) NS3 protease inhibitors[6] and another series of acetylcholinesterase (AChE) inhibitors,[7] have been chosen because they are representative cases of lead optimization, in which the compounds share a common scaffold. In the first series, there is only one site of chemical substitution, while in the second there are two. The third data set contains a wide series of inhibitors of factor Xa extracted from the

* Corresponding author. Tel: +34 932 240 302. Fax: +34 932 240 875. E-mail: mpastor@imim.es.

10.1021/jm049113+ CCC: $30.25
Published on Web 03/02/2005

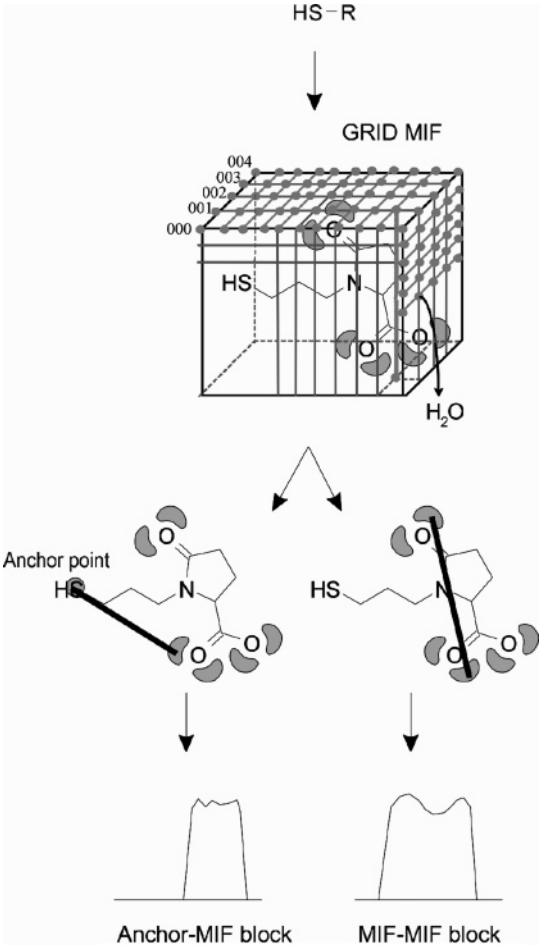


Figure 1. Calculation of the anchor−GRIND descriptors for an ACE inhibitor with the anchor point set on the zinc binder sulfur atom.

literature, all sharing a benzamidine moiety. The amidine group is a clear anchoring point that is conserved in a lot of factor Xa inhibitors and was therefore used as such in the present study.

Material and Methods

Most of the series involved in lead optimization processes are constituted by compounds that can be represented by means of a Markush structure, with the chemical groups susceptible of modifications referred as R groups. Such series are characterized by sharing a structural scaffold, common to all the compounds of the series, that, provided that the interactions between the scaffold and the binding site are specific enough, defines a pattern of interactions shared by all the compounds in the series. In the Markush structure, the R substituents refer to the structural parts that are specific to one compound or a subset of the compounds present in the series.

In the anchor−GRIND approach the R groups are described with two blocks of variables: the anchor−MIF and the MIF−MIF blocks (Figure 1). The first one describes the geometrical distribution of the R MIF relative to the anchor point, while the second one describes the geometrical distribution of the MIF within each R group. These blocks are obtained after the following steps: (i) every R group is considered as attached to the scaffold, (ii) the anchor point is set automatically on an atom of the scaffold, (iii) a set of MIF are calculated with the program GRID as well as the shape field implemented in the program Almond,[8] and (iv) as the last step, the blocks of descriptors are computed from the anchor point and the filtered MIF.

Structures and Anchor Point. In the example series, the R groups and their scaffold were written in SMILES[9] and converted to a valid 3D structure with CORINA.[10] The anchor point atom was defined by means of the SMART[11] chemical language. The position of the anchor point in the 3D space was assigned automatically using a program written ad hoc, using the Open Babel library.[12]

Fields Computations and Filtering. The program GRID 21[13] was used to compute the MIF of three probes: the hydrophobic probe DRY, the hydrogen-bond-acceptor probe O, and the hydrogen-bond-donor probe N1. The program Almond[8] was used to compute a shape field[14] with the probe TIP.

The MIF interactions in which the main contributor was a scaffold atom were filtered out, except those that are situated in the vicinity of the R substituents. Indeed, the interactions generated by the scaffold are the same for all the compounds of the series, and keeping the interactions of the scaffold would generate variables with exactly the same value for all the compounds, which are therefore useless. For QSAR, the removal of the scaffold interactions simplifies the description and prevents the overlapping of distances measured within the scaffold with other distances measured between variable parts of the structures and which might be more relevant.

Descriptors Computation. The descriptors computation was performed with an in-house version of the software Almond 3.3a, using default parameters and the anchor point options activated. The number of R substituents on the scaffold determines the number of blocks of descriptors required for the analysis. In the simplest case, there is only one site of chemical modifications (e.g., HCV NS3 data set) and two blocks of variables. The first block is used to encode the geometrical relationships between the anchor point and the MIF (Figure 1) and is particularly relevant to characterize the substituents structure in a simple way. The second block of variables is used to describe the geometrical relationships between the MIF of the substituents (Figure 1). This is particularly useful to discriminate different pharmacophoric features that are located at the same distance from the anchor point. In more complex cases of lead optimization in which the scaffold holds more than one substituent (e.g., AChE data set), there are as many blocks as R groups plus an additional one that encodes inter- and intra-R group MIF features.

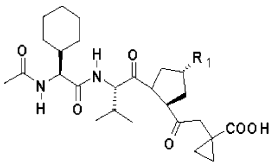
Results and Discussion

Three data sets have been selected to illustrate the applicability of the anchor−GRIND methodology: the first one contains 32 HCV NS3 inhibitors, the second one 49 AChE inhibitors, and the third one contains 435 factor Xa inhibitors compiled from the literature.

HCV Data Set. The hepatitis C virus NS3 protease is a serine protease that is essential for the viral infection and therefore is a widely studied and validated target for drug design. The structure of the 32 inhibitors considered and their IC_{50} are shown in Table 1.[6] The IC_{50} spans 2.5 log units. This is a typical case of lead optimization, in which structure−activity models would be useful for guiding the synthesis of new compounds. The anchor point chosen for the model is the carboxy terminal group, which is common to all the peptidic inhibitors. This functional group is crucial for the activity[15] and, as revealed by crystallographic studies, forms a hydrogen bond with the catalytic serine nitrogen.[16]

As described above, the molecular descriptors were divided in two blocks: the anchor−MIF and the MIF−MIF blocks. The two blocks of descriptors were joined into a single matrix of descriptors and statistically analyzed with the GOLPE[17] program. Principal component analysis (PCA) on this matrix showed that the reagent intermediate **10** is an outlier because of the very

Table 1. Inhibitors of the Hepatitis C Virus NS3 Protease



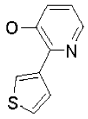
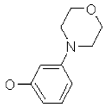
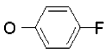
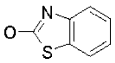
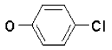
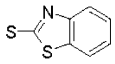
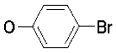
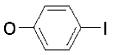
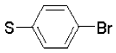
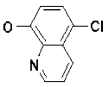
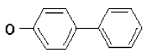
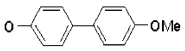
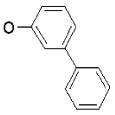
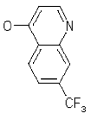
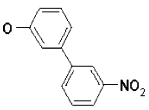
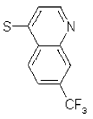
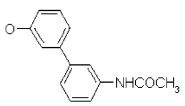
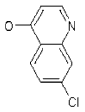
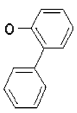
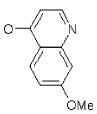
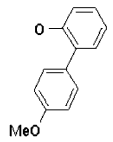
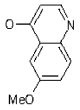
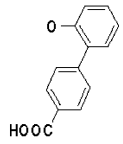
Compound	R_1	log $1/IC_{50}$	Compound	R_1	log $1/IC_{50}$
10	OH	<3	**27**		<3.7
11		4.1	**28**		4.5
12		4	**29**		4.7
13		4.3	**30**		4.7
14		4.6	**31**		3.8
15		5	**32**		3.7
16		4.4	**33**		3.6
17		4.3	**34**		4
18		4.6	**35**		5.1
19		4.5	**36**		5.4
20		4.5	**37**		5.3
21		5.2	**38**		5.2
22		5.2	**39**		5.0
23		5.2	**40**		5.7
24		<3.7	**41**		6.1

Table 1 (Continued)

Compound	R_1	log $1/IC_{50}$	Compound	R_1	log $1/IC_{50}$
25		<3.7	**42**		4.8
26		<3.7			

small size of the substituent, and therefore, compound **10** was excluded from the partial least-squares (PLS) regression. The initial PLS model incorporating all the variables of the anchor−MIF and the MIF−MIF blocks showed better predictive ability (3 LV, $r^2 = 0.86$, $q^2 = 0.45$) than the PLS model with only the variables of the anchor−MIF block (3 LV, $r^2 = 0.76$, $q^2 = 0.35$). Normally, if both models produce similar statistical parameters, only the model with the anchor−MIF block is kept, since it is simpler to analyze. However, the higher predictive ability of the two-block model means that the MIF−MIF block contains some additional relevant structural information that is not present in the anchor−MIF block. Details about these structural features are given below. After manual exclusion of irrelevant variables, the predictive ability of the model increased (3 LV, $r^2 = 0.86$, and $q^2 = 0.52$) as indicated by the q^2 value. Automatic variable selection by means of fractional factorial design (FFD) improved considerably the model quality (3 LV, $r^2 = 0.88$, $q^2 = 0.72$). Variables anchor−DRY 12.8 Å, anchor−N1 6.8 Å, and anchor−TIP 17.6 Å were identified as the most important anchor variables with a positive weight on the model. Conversely, variables anchor−DRY 14.4 Å and anchor−TIP 13.2 Å were the anchor variables with the most negative weight on the model. The interpretation of each important variable is straightforward, because the regions of the MIF corresponding to these variables can be visualized with the 3D plot application included in Almond. Taken together, the anchor−DRY and the anchor−TIP variables define the ideal size and shape of a hydrophobic cavity (Figure 2). This cavity is more appropriate for accommodating two fused aromatic cycles (e.g. compound **41**) than a single one (e.g. compound **32**). Moreover, the cavity does not seem to accommodate biphenyl with the second phenyl group ortho to the first one, as indicated by the variable anchor−DRY 14.4 Å. The most important variables of those having positive weight from the MIF−MIF block are the N1−N1 10.4 Å and N1−TIP 11.2 Å ones, while those with the most negative weight are DRY−DRY 10 Å, N1−N1 4 Å, and N1−TIP 3.2 Å. Variable DRY−DRY 10 Å and anchor−DRY 14.4 Å supply approximately the same information, i.e., a biphenyl with the second phenyl group ortho to the first one is not well tolerated by the cavity, although the correlation between these two variables is low ($r^2 = 0.4$). Taken together, variables anchor−N1 6.8 Å, N1−N1 10.4 Å, and N1−TIP 11.2 Å indicate the ideal position of a hydrogen-bond-donor site caused by an aromatic nitrogen (e.g. compound **41**). Variable anchor−N1 6.8 Å shows the ideal distance between the hydrogen-bond-donor site and the carboxyl group (Figure 3a), N1−N1 10.4 Å indicates the ideal distance between the hydrogen-bond-donor site and the phenoxy oxygen (Figure 3b), and finally, the N1−TIP 11.2 Å variable shows the ideal distance between the hydrogen-bond-donor site and the end of the cavity (Figure 3c). In conclusion, the anchor−GRIND approach produces satisfying results both in terms of statistical parameters and interpretation. For comparison, the standard GRIND method produces a model with a lower q^2 (3 LV, $r^2 = 0.71$, $q^2 = 0.52$) and is much more difficult to interpret. The superiority of the anchor−GRIND model over the standard GRIND is due to a more specific description, since the interactions represented by the variables are more unique and less likely to be confounded with interactions happening at similar distances but involv-

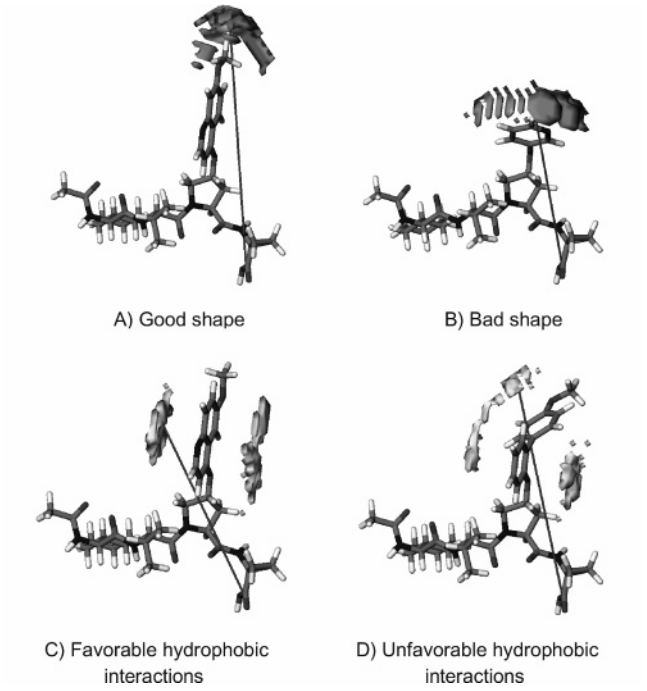


Figure 2. Anchor−MIF variables defining the ideal shape of the hydrophobic cavity for the HCV inhibitors: (A) variable anchor−TIP 17.6 Å, compound **41**; (B) variable anchor−TIP 13.2 Å, compound **32**; (C) variable anchor−DRY 12.8 Å, compound **41**; (D) variable anchor−DRY 14.4 Å, compound **25**.

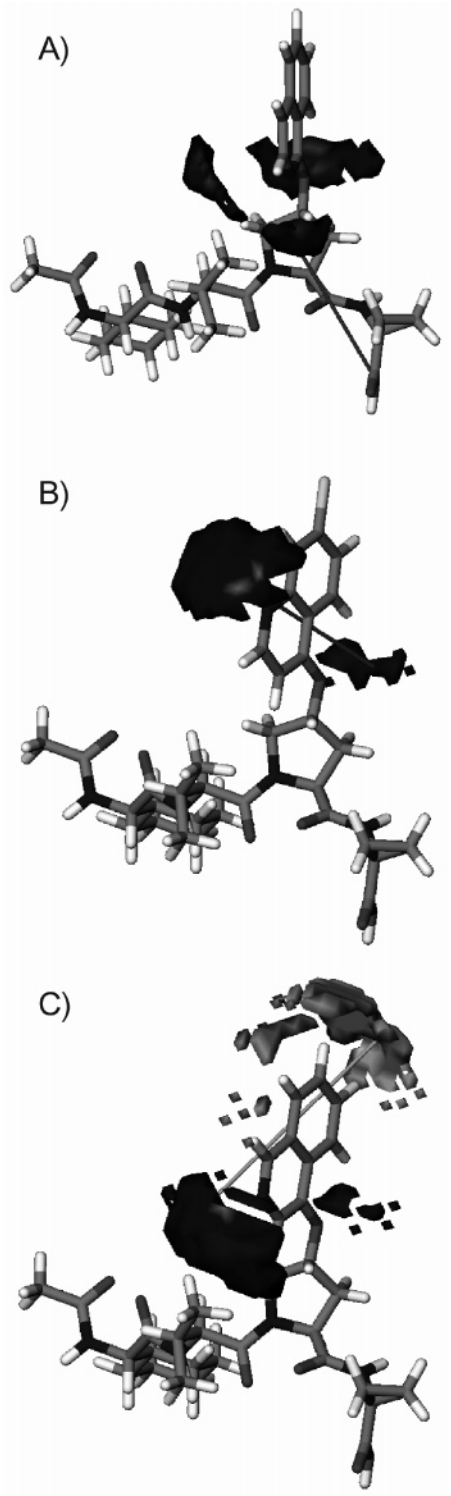
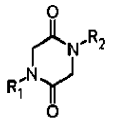


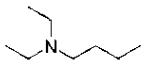
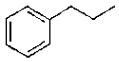
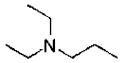
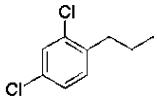
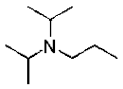
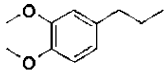
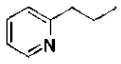
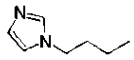
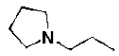
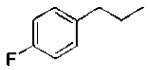
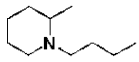
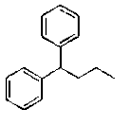
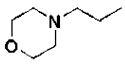
Figure 3. Variables that define the ideal position of a hydrogen-bond-donor site for the HCV inhibitors: (A) anchor−N1 6.8 Å, compound **35**; (B) N1−N1 10.4 Å, compound **40**; (C) N1−TIP 11.2 Å, compound **36**; N1, black; TIP, gray.

ing only atoms of the common scaffold. Indeed, as expected, the particular requirement of the anchor−GRIND methodology (the existence of a common anchor point or scaffold for all the compounds) limits its applicability but produces a relevant improvement in the quality of the models when it can be used.

AChE Data Set. The series of 2,5-piperazinedione (DKP) is shown in Table 2.[7] These compounds differ in substituents in two positions (R_1 and R_2). R_1 groups are

Table 2. AChE Inhibitors



	R_1		R_2
a		1	
b		2	
c		3	
d		4	
e		5	
f		6	
g		7	

essentially aryl substituents, while R_2 are essentially tertiary amines. All the active compounds have an IC_{50} between 2 and 20 μM, while compounds with an IC_{50} above 20 μM were classified as inactive. The anchor point was set on one of the carbonyl oxygens of the scaffold, which interacts with a crystallographic water molecule of the binding site according to molecular modeling studies.[7] As a first analysis of the data, a PCA has been performed using only the descriptors of the R_1 and R_2 blocks. Due to the particular design of the series, such blocks of descriptors are not intercorrelated and, consequently, the first PC describes only the changes of R_1, while the second PC just describes the changes of R_2 (Figure 4). Consequently, in the score plot, the compounds are distributed as if they were in a table, which facilitates the analysis. Interestingly, the active compounds are clustered on the top of the score plot, indicating that the similarity principle is respected for this data set, i.e., similar compounds have similar activity. Accordingly, a chemical space containing the seven hits of the assay and only three inactives can be defined in the PCA score plot. The structure of the three inactive compounds is very similar to the structure of some active compounds and probably their activity is not far from the active/inactive cutoff. Unfortunately, we cannot verify this hypothesis, since the activities of the inactive compounds have not been published.

In the following step, PLS discriminant analysis was carried out in order to identify the structural features that differentiate active from inactive compounds. Two

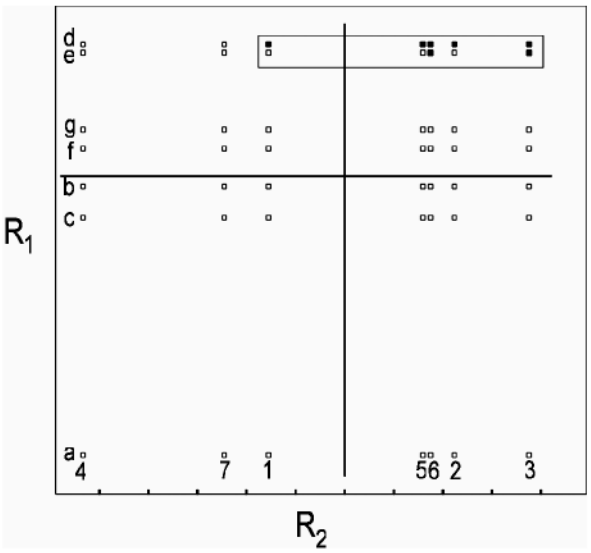


Figure 4. PCA scores plot for the DKP library. PC1 explains variation of the R_2 group, whereas PC2 explains variation of the R_1 group. Black squares represent active compounds. The upper square encloses a subspace containing the most active compounds.

configurations were tried, one with only the R_1 and R_2 anchor−MIF blocks and another one adding the MIF−MIF block as explained in the previous example. Both produced the same discriminant efficacy, i.e., 92% good classification in cross-validation. Therefore, only the first model including only the anchor−MIF blocks will be described hereafter, since it is simpler to analyze.

Four anchor−MIF variables important for the activity, three on the R_1 block and one on the R_2 block, have been found (Figure 5). Two anchor−TIP variables (anchor−TIP 8.4 Å and anchor−TIP 12.4 Å) refer to the favorable shape of the methoxy groups of substituent d (Figure 5a). Anchor−N1 9.6 Å corresponds to the favorable hydrogen-bond-acceptor capability of substituents d and e (Figure 5b). The remaining favorable variable is anchor−O 7.6 Å (Figure 5c), which represents the ideal distance between the anchor point and the hydrogen-bond-acceptor site generated by the protonated nitrogen. In summary, at the R_1 position active compounds have an aromatic moiety bearing a hydrogen-bond-acceptor atom and connected to the scaffold by a linker of two or three carbon atoms, whereas at R_2 position active compounds contain a tertiary amine connected to the scaffold by a linker of two or three atoms.

Factor Xa Data Set. The objective of this example was to perform a retrospective 3D-QSAR modeling of the published inhibitors of factor Xa of the benzamidine family. In this series, the reason for selecting the anchor point is not chemical but biological: the anchor point chosen was the amidine group, a functional group present in many factor Xa inhibitors and which binds to the factor Xa in a well-defined pocket. The considered data set contains 156 low-activity compounds (K_i higher than 1 μM) and 279 very active compounds (K_i lower than 10 nM). PLS discriminant analysis was used in order to separate the two groups of compounds. Since the data set contains compounds of very diverse chemical classes, the objective is not to give a detailed picture of the interactions made by each compound studied but

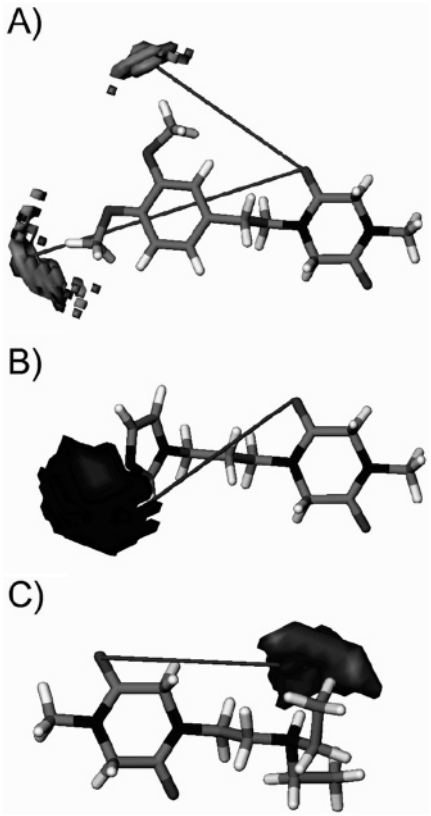


Figure 5. Favorable anchor−MIF variables for the DKP library: (A) R_1 group, anchor−TIP 8.4 Å and anchor−TIP 12.4 Å; (B) R_1 group anchor−N1 9.6 Å; (C) R_2 group, anchor−O 7.6 Å.

to produce a discriminant model, potentially useful for the screening of molecular libraries. To validate the discriminant analysis, the data set has been split in a training (290 compounds) and a test set (145 compounds) using random assignment. Two models were tested, one with two blocks of variables, i.e., the anchor−MIF and the MIF−MIF block, and one with only one block of variables, i.e., the anchor−MIF block. The results for the test set are slightly better for the one-block model: 88% of the compounds are well classified versus 84% of the compounds being well classified in the two-block model. This slight difference in the quality of the prediction indicates that the MIF−MIF block does not add relevant structural information that is not already present in the anchor−MIF block. The slight decrease of correct predictions for the two-block model is probably due to the fact that the variables of the MIF−MIF block are adding only noise to the model. The one-block model is simpler than the two-block model, since it includes fewer original variables (185 versus 742). Consequently, the one-block model was considered for the subsequent analyses.

The PLS coefficients profile and the PLS scores plot for two LV are shown in Figure 6. The three variables with the most positive weight on the model are anchor−DRY 12 Å, anchor−O 7.6 Å, and anchor−N1 4.4 Å (Figure 7). Anchor−DRY 12 Å defines the ideal position of a hydrophobic region interacting with an aromatic ring. Anchor−O 7.6 Å often corresponds to the presence of a hydroxyl group in the para position of the benzamidine ring (Figure 7b). Anchor−N1 4.4 Å indicates the ideal distance between the amidine group and a hydro-

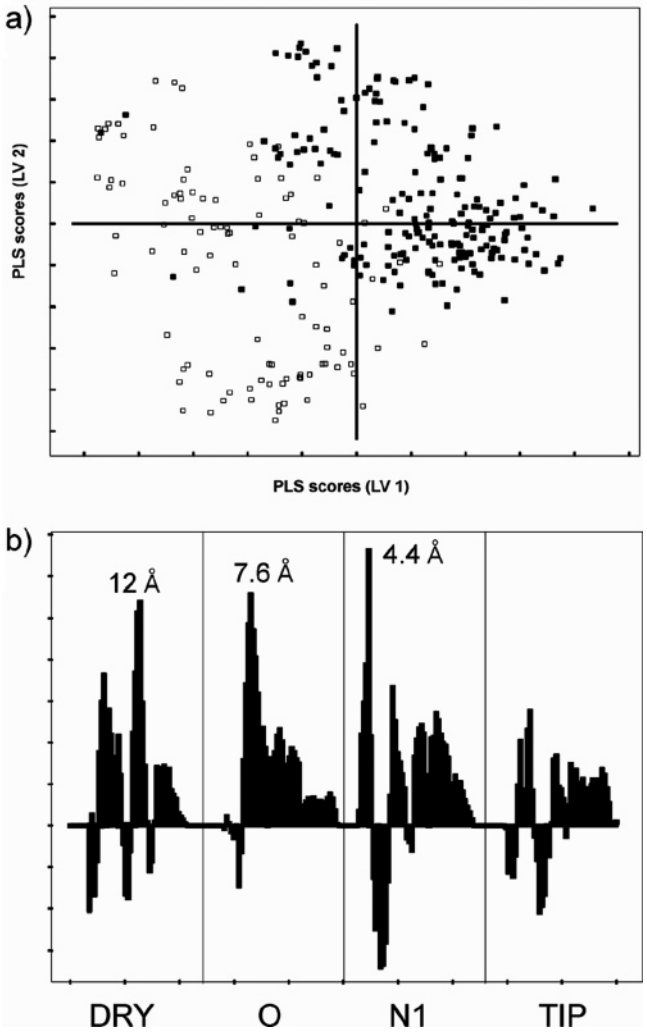


Figure 6. (A) Two LV PLS score plot for the factor Xa model: black squares, compounds with high activity; white squares, compounds with low activity. (B) two LV PLS coefficients.

gen-bond-donor site. It should be emphasized that the distances indicated here, although relevant for the discriminant analysis, might not correspond to actual distances between interaction sites of the protein, since the conformation of the compounds were a straight output from CORINA. To check this issue, we examined the crystal structure of factor Xa complexed with the highly active inhibitor RPR128515 (PDB code 1EZQ). Apparently, every distance obtained in the anchor−GRIND analysis corresponds to a particular residue of the binding site. Anchor−DRY 12 Å might correspond to residues Phe174 and Tyr99, which have side chain atoms sandwiching a phenyl of the inhibitor at a distance between 12 and 11 Å, respectively. Anchor−O 7.6 Å might correspond to the hydroxyl group of Ser195, which is at 7.2 Å from the amidine carbon atom. Finally, anchor−N1 4.4 Å seems to refer to the amide nitrogen of Gly216, which is at 4.0 Å from the amidine carbon atom. These good correlations between the distances and the anchor−GRIND variables are probably due to the fact that the shape of the factor Xa binding cavity forces the inhibitors to bind in a fairly extended conformation similar to the one obtained from CORINA.

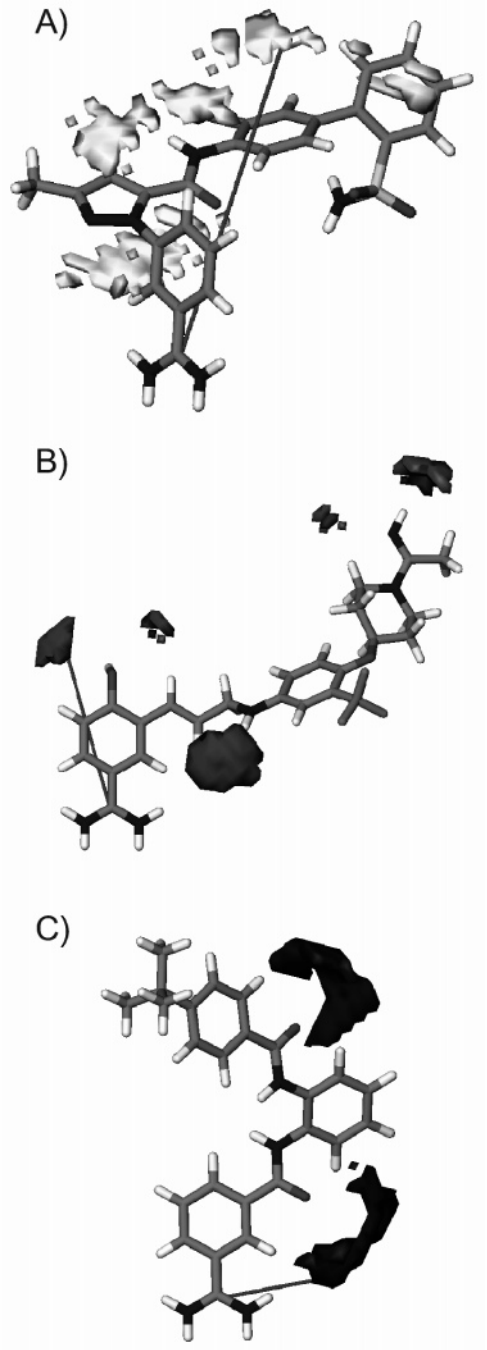


Figure 7. Some highly active compounds with the variables having the most positive weight in the factor Xa model: (A) anchor−DRY 12 Å, (B) anchor−O 7.6 Å; (C) anchor−N1 4.4 Å.

Conclusions

In the present study, we have introduced the anchor−GRIND method. This method represents a particular application of the GRIND for situations where there is at least a single point in the structure of the considered compounds that can be recognized as common for either chemical or biological reasons. The descriptors obtained are far more specific and therefore produce better models, which are easier to interpret. In addition, the interactions within atoms of the common scaffold are filtered out so that the information contained in the descriptors is highly specific for the sites of variability.

The improvement obtained with this new methodology has been demonstrated with three different ex-

amples, in which the novel approach allowed to identify in a rather straightforward way the most relevant structural features of active and inactive compounds. The results for the factor Xa data set are particularly interesting, since they show that, when a relevant anchor point can be defined, excellent models can be obtained even for highly diverse compounds. It is also impressive to see how the anchor−GRIND distances match structural features of the factor Xa binding site, since this information was not considered for obtaining the model.

Acknowledgment. This work has been supported by funding from the "Fundació La Marató TV3".

References

(1) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA): 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.

(2) Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. GRid-INdependent descriptors (GRIND): A novel class of alignment-independent three-dimensional molecular descriptors. *J. Med. Chem.* **2000**, *43*, 3233−3243.

(3) Cruciani, G.; Pastor, M.; Guba, W. VolSurf: A new tool for the pharmacokinetic optimization of lead compounds. *Eur. J. Pharm. Sci.* **2000**, *11 Suppl 2*, S29−39.

(4) Mason, J. S.; Morize, I.; Menard, P. R.; Cheney, D. L.; Hulme, C.; et al. New 4-point pharmacophore method for molecular similarity and diversity applications: Overview of the method and applications, including a novel approach to the design of combinatorial libraries containing privileged substructures. *J. Med. Chem.* **1999**, *42*, 3251−3264.

(5) Zamora, I.; Afzelius, L.; Cruciani, G. Predicting drug metabolism: A site of metabolism prediction tool applied to the cytochrome P450 2C9. *J. Med. Chem.* **2003**, *46*, 2313−2324.

(6) Poupart, M. A.; Cameron, D. R.; Chabot, C.; Ghiro, E.; Goudreau, N.; et al. Solid-phase synthesis of peptidomimetic inhibitors for the hepatitis C virus NS3 protease. *J. Org. Chem.* **2001**, *66*, 4743−4751.

(7) Carbonell, T.; Masip, I.; Sanchez-Baeza, F.; Delgado, M.; Araya, E.; et al. Identification of selective inhibitors of acetylcholinesterase from a combinatorial library of 2,5-piperazinediones. *Mol. Divers.* **2000**, *5*, 131−143.

(8) Cruciani, G.; Fontaine, F.; Pastor, M. *Almond*; 3.3.0; Molecular Discovery Ltd.: Perugia, Italy, 2004.

(9) Weininger, D. Smiles, a Chemical Language and Information-System 0.1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31−36.

(10) Gasteiger, J.; Rudolph, C.; Sadowski, J. Automatic generation of 3D atomic coordinates for organic molecules. *Tetrahedron Comput. Methodol.* **1990**, *3*, 537−547.

(11) http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html.

(12) http://openbabel.sourceforge.net/.

(13) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.

(14) Fontaine, F.; Pastor, M.; Sanz, F. Incorporating molecular shape into the alignment-free GRid-INdependent Descriptors. *J. Med. Chem.* **2004**, *47*, 2805−2815.

(15) Llinas-Brunet, M.; Bailey, M.; Deziel, R.; Fazal, G.; Gorys, V.; et al. Studies on the C-terminal of hexapeptide inhibitors of the hepatitis C virus serine protease. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2719−2724.

(16) Di Marco, S.; Rizzi, M.; Volpari, C.; Walsh, M. A.; Narjes, F.; et al. Inhibition of the hepatitis C virus NS3/4A protease. The crystal structures of two protease-inhibitor complexes. *J. Biol. Chem.* **2000**, *275*, 7152−7157.

(17) Pastor, M.; Barroni, M. *GOLPE*; 4.5.11; Multivariate Infometric Analysis: Perugia, Italy, 2005.

JM049113+