J Comput Aided Mol Des (2007) 21:63–86
DOI 10.1007/s10822-007-9106-2

ORIGINAL PAPER

Outliers in SAR and QSAR: Is unusual binding mode a possible source of outliers?

Ki Hwan Kim

Received: 17 July 2006 / Accepted: 9 January 2007 / Published online: 3 March 2007
© Springer Science+Business Media B.V. 2007

Abstract A lead optimization is usually carried out by structure-activity relationship (SAR) and/or quantitative structure-activity relationship (QSAR) studies. One of the assumptions in SAR and QSAR studies is that similar analogs bind to the same binding site in a similar binding mode. One often observes that there are outliers, especially in QSAR. However, most QSAR studies are carried out focusing their attention to the development of QSAR and leave the outliers without much attention. We searched a number of ligand-bound X-ray crystal structures from the protein structure database to find evidences that could indicate a possible source of outliers in SAR or QSAR. Our results show that unusual binding mode could be a source of outliers.

Keywords SAR · QSAR · Outlier · Binding mode · X-ray structures · Source of outliers

Introduction

Structure-activity relationship (SAR) study has long been playing an important role in drug discovery research for optimizing a lead compound. A computational method describing SAR quantitatively is called quantitative structure-activity relationship (QSAR), and the method has been used over 40 years. Hansch and co-workers have collected physical and biological QSAR equations and created the world largest QSAR database called C-QSAR. The QSAR database now contains more than 18,000 biological equations, including many unpublished ones [1]. The database is available through the BioByte Corporation with the Bio-Loom interface [2].

It has often been observed that there are outliers in SAR or QSAR. One finds that there are a significant number of cases where one or more compounds were left out to develop the reported QSARs in the C-QSAR database. Outliers of QSARs can be very important and interesting especially when the observed biological activity is higher than the predicted one by the QSAR. Of course, an outlier may point toward an experimental or even a typographical error. They may also imply that the QSAR lacks certain descriptors or parameters to describe the QSAR for the entire compounds in the dataset, or the mathematical model is not appropriate. Outliers may be due to the inappropriate calculation of the parameter values used. They may indicate a different mode of mechanism. Outliers may also result from a different binding mode.

Although many researchers have studied and reported QSARs of various sets of compounds, they usually focused their attention to the development of QSAR and mostly left the outliers without much discussion. To the knowledge of this author, no study has yet focused on the subject of outliers in SAR or QSAR.

In this study, we focus our initial efforts on finding structural evidences that might indicate a possible source of outliers in SAR or QSAR. Outliers may suggest a different direction for the lead optimization.

Electronic supplementary material The online version of this article (doi:10.1007/s10822-007-9106-2) contains supplementary material, which is available to authorized users.

K. H. Kim (✉)
Hope Drug Discovery Research Laboratory, 260 Southgate Drive, Vernon Hills, IL 60061, USA
e-mail: pkhkim@gmail.com

Experimental section

C-QSAR database [1] searching

The database was searched with a query 'carbonic anhydrase' using the BioByte Corporation's Bio-Loom interface [2]. The resulting list of the search was further examined for individual QSAR summary for the number of outliers, the number of outliers with the deviation greater than three times of its standard deviation, and the number of outliers with the deviation grater than 1.0 or 2.0 positive and negative deviations. The percentage of the outliers versus the total number of compounds in each dataset and the average percentage of the outliers from the entire datasets were then calculated.

RCSB protein data bank [3] searching

The database was searched for most of the compounds presented in this paper based on the literature information about unusual binding modes. For the DHFR inhibitors, protein-ligand database searching program Relibase (version 1.3.2) [4] was used. The resulting pdb files from the searches were then further examined for their binding modes by superimposing the protein structures in each sub-group. The multiple sequence alignment for protein structure comparison was done using the ClustalW program of the EMBL-EBI [5].

Molecular graphics

All the figures were generated using the UCSF Chimera molecular modeling program (beta version 1) [6] using the multiple sequence alignments obtained from ClustalW described above.

Results and discussion

Outliers in QSAR equations

In order to discover possible sources of outliers in QSAR, we first searched the C-QSAR database for all the reported QSAR equations using a particular search query to examine how frequently outliers occur in QSAR. The search query used as a test case was 'carbonic anhydrase'. Sixty equations resulted from the search. They are listed in Table 1 (See the Supplemental Material for the entire list). It was interesting to find that only six out of the 60 QSAR equations found for the 'carbonic anhydrase' query utilized all compounds available for the QSAR studies. The other 54 equations have one or more outliers. The number of equations with outliers is more than what we had initially expected (Fig. 1, See the discussion below.). The biological activities for these QSARs are the inhibitory activity (log 1/C) or Ki inhibition of the enzyme (log 1/Ki, log KI). In general, the quality of the reported equations is very good judged by their standard deviations and squared correlation coefficients. Many of these correlations can also be supported by the lateral validation [7, 8].

For the 54 'carbonic anhydrase' QSARs with one or more outliers, the percentage of the outliers with respect to the total number of compounds ranges from 2% up to 33%. Thirty of the 54 datasets have more than 20% of outliers and the average percentage of the outliers from the entire 60 datasets is 21.7%. This means that on the average two out of every 10 compounds in each dataset do not fit to the QSAR model described. Such a high percentage of outliers are surprising, and it would be nice if one can offer some explanations.

From further analysis of the QSARs in Table 1, we found that among the 155 compounds left out for QSARs in the entire sets of 'carbonic anhydrase' QSARs, 107 outliers have the standard deviations greater than three-times of the standard deviation of the corresponding correlations. The other 48 have the standard deviations less than three fold of the standard deviation of the corresponding equation, but they were treated as outliers. The number of compounds with a positive deviation (compounds calculated to be less potent than the observed) and the number of compounds with a negative deviation (compounds calculated to be more potent than the observed) are not much different: 59 vs. 48. Thirty of the 107 compounds have the standard deviations greater than one-logarithm unit (10-fold deviation), and 16 of the 107 have the standard deviations greater than two-logarithm unit (100-fold deviation). For these outliers, the number of compounds with a positive deviation and the number of compounds with a negative deviation are also not much different: 16 vs. 14, and 7 vs. 9, respectively.

Examination of the outliers revealed some unusual cases. In particular, two equations in Table 1 need special comments. The first is C-QSAR equation number 12512 (No. 55 in Table 1). One outlier in this dataset has the deviation of 7.25 in log 1/KI for the correlation derived from only four compounds. The observed biological activity of this compound is 1 million times less potent than the observed. Another is C-QSAR equation number 12397 (No. 49 in Table 1). All five outliers in this set are calculated to be 100-times more potent than the experimentally measured activity value. Their deviations are much larger than normal in magnitude, and all have a large

Table 1 QSAR equations for carbonic anhydrase searched from C-QSAR database. For the all 60 equations, see the supplemental material

	Eq. No.[a]	n[b]	Out[c]	%	+; –[d]	+; –[e]	+; –[f]	s.d.[g]	R^{2h}	Opt[i]		Equation
49	12397	16	5	24	0; 5	–	0; 5	0.513	0.863	8.949	Log $^1/_{KI}$ =	0.80·CMR – 2.38·BILIN(CMR) + 1.90
55	12512	4	1	20	1; 0	–	1; 0	0.315	0.974		Log $^1/_{KI}$ =	0.19·NVE – 15.63

[a] Equation number in C-QSAR database
[b] Number of compounds included in the correlation
[c] Number of outliers
[d] Number of outliers with the deviation greater than three times of s.d. Positive deviation; Negative deviation
[e] Number of outliers with the deviation greater than 1.0. Positive deviation; Negative deviation
[f] Number of outliers with the deviation greater than 2.0. Positive deviation; Negative deviation
[g] Standard deviation
[h] Squared correlation coefficient
[i] Optimum or minimum value of the parabolic or inversed parabolic equation

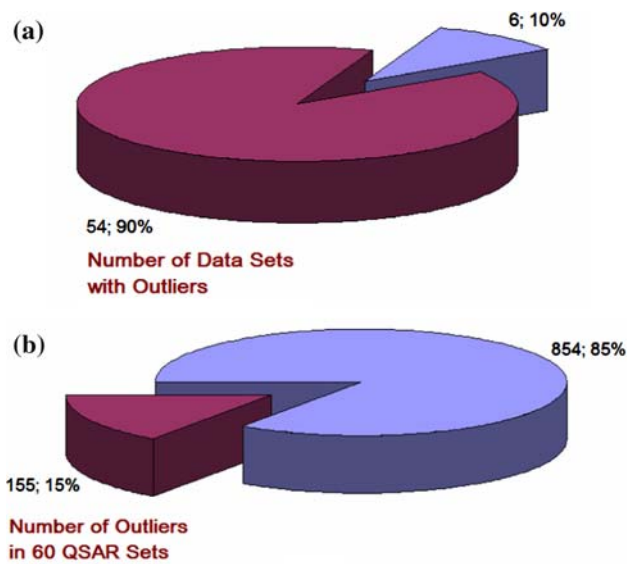


Fig. 1 (**a**)The number of data sets with outliers (54; 90%) and without outliers (6; 10%). Ninety percent of the 60 QSAR data sets are with outliers. (**b**) The number of compounds included (854; 85%) and outliers (155; 15%) in the 60 carbonic anhydrase inhibitor QSAR sets. Fifteen percent of the compounds in the 60 QSAR data sets are outliers

negative deviation (–2.02 ~ –3.62 in log 1/KI). The standard deviation itself for this correlation is also unusually high (0.513).

What are the possible sources of outliers in QSAR (or SAR)? We attempt to search a possible source from the binding modes of ligands suggested by the crystal structure data.

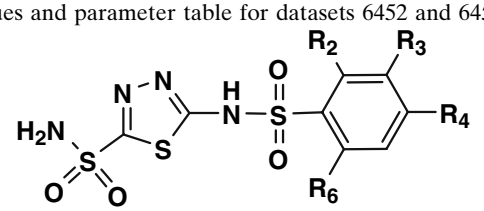
A possible source of outliers in SAR and QSAR: Cases of unusual binding modes

In SAR or QSAR, one usually first assumes that analogous molecules of interest bind to the same binding site in an essentially identical binding mode. Let us look at specific examples. Human cloned carbonic anhydrase I (HCA-I) or carbonic anhydrase II (HCA-II) inhibitors of 5-benzenesulfonylamino[1,3,4]thiadiazole-2-sulfonic acid amide (Structural type I, Table 2) analogs were assumed to bind to the same binding site of HCA-I or HCA-II in a same binding mode. Their QSARs can be found in C-QSAR database[2]: datasets 6452 and 6453.

The QSAR results for the data set 6452 are as follows and summarized in Table 2: log $^1/_C = 0.16\cdot\sigma^+ - 0.77\cdot$B1-2 – 0.20·B5-4+9.56, N = 11, s.d. = 0.093, $R^2 = 0.920$ (4 outliers), where C is the inhibitory concentration (IC_{50}) of these compounds against HCA-I. The QSAR indicates that the inhibitory potency of these compounds increases as the sum of the calculated electronic properties of the phenyl ring substituents expressed as σ^+ increases. The activity decreases as the steric property of R_2 expressed as B1-2 and that of R_4 expressed as B5-4 increase. No other examined descriptors improved the quality of the correlation already obtained. Four compounds were not included in developing the reported QSAR equation: 2-COOH, 3-NH_2, 4-OMe, and 3-NO_2-4-Cl. The 2-COOH analog may not fit to the QSAR and could be left out, because it is the only negatively charged molecule in this data set. It has the largest deviation. However, there appears to be no obvious reason why the other three analogs (3-NH_2, 4-OMe, and 3-NO_2-4-Cl) do not fit to the QSAR obtained.

The QSAR results for the data set 6453 are as follows and summarized in Table 2: log $^1/_C = -0.21\cdot$Clog$P$ – 0.54·B1-2 – 0.18·B5-4 + 9.53, N = 11, s.d. = 0.111, $R^2 = 0.895$ (4 outliers), where C is the inhibitory potency (IC_{50}) of these compounds against HCA-II. The results indicate that the inhibitory

Table 2 Observed, predicted log 1/C values and parameter table for datasets 6452 and 6453 [2]



(I)

6452: log $^1/_C = 0.16 \cdot \sigma^+ - 0.77 \cdot$B1-2 – 0.20·B5-4 + 9.56, N = 11, s.d. = 0.093, $R^2 = 0.920$ (4 outliers), 6453: log $^1/_C = -0.21 \cdot$Clog*P* – 0.54·B1-2 – 0.18·B5-4 + 9.53, N = 11, s.d. = 0.111, $R^2 = 0.895$ (4 outliers)

No	Compounds	HCA-1 log $^1/_C$ (6452)			σ^+	B1-2	B5-4	HCA-II log $^1/_C$ (6453)			Clog*P*	B1-2	B5-4
		Obs	Pred	Dev				Obs	Pred	Dev			
2-1	4-Me	8.30	8.34	–0.04	–0.31	1.00	2.04	8.40	8.52	–0.12	0.48	1.00	2.04
2-2	4-F	8.40	8.52	–0.12	–0.07	1.00	1.35	8.40	8.72*	–0.32	0.14	1.00	1.35
2-3	4-Cl	8.40	8.45	–0.06	0.11	1.00	1.80	8.52	8.52	0.00	0.71	1.00	1.80
2-4	4-Br	8.52	8.43	0.09	0.15	1.00	1.95	8.70	8.46*	0.24	0.86	1.00	1.95
2-5	4-OMe	8.30	8.06*	0.24	–0.78	1.00	3.07	8.52	8.42	0.10	0.08	1.00	3.07
2-6	4-NHCOMe	8.00	7.98	0.02	–0.60	1.00	3.61	8.52	8.48	0.04	–0.68	1.00	3.61
2-7	4-NH_2	8.22	8.20	0.02	–1.30	1.00	1.97	8.70	8.83	–0.13	–0.90	1.00	1.97
2-8	3-NH_2	8.05	8.57*	–0.53	–0.16	1.00	1.00	9.00	9.00	0.00	–0.90	1.00	1.00
2-9	4-NO_2	8.52	8.43	0.09	0.79	1.00	2.44	9.00	8.60*	0.40	–0.22	1.00	2.44
2-10	3-NO_2	8.70	8.71	–0.01	0.71	1.00	1.00	9.05	8.86	0.19	–0.22	1.00	1.00
2-11	2-NO_2	8.30	8.19	0.12	0.79	1.70	1.00	8.52	8.52	0.00	–0.40	1.70	1.00
2-12	2-COOH	9.00	8.20*	0.80	0.42	1.60	1.00	9.30	8.52*	0.78	–0.14	1.60	1.00
2-13	2,4-$(NO_2)_2$	7.92	8.02	–0.10	1.58	1.70	2.44	8.30	8.31	–0.01	–0.64	1.70	2.44
2-14	3-NO_2,4-Cl	8.05	8.57*	–0.52	0.82	1.00	1.80	8.52	8.60	–0.08	0.30	1.00	1.80
2-15	2,4,6-$(Me)_3$	7.82	7.84	–0.02	–0.93	1.52	2.04	8.05	8.03	0.01	1.48	1.52	2.04

* Not included in deriving QSAR

potency of these compounds decreases as the sum of the calculated lipophilicity of the phenyl ring substituents expressed as Clog*P* increases. The activity also decreases as the steric property of R_2 expressed as B1-2 and that of R_4 expressed as B5-4 increase. Similar to the QSAR mentioned earlier, no other examined descriptors improved the quality of the correlation already obtained. Four compounds were not used in developing the reported QSAR equation: 2-COOH, 4-F, 4-Br, and 4-NO_2. Again, the 2-COOH analog may not fit to the QSAR and could be left out because it is the only negatively charged molecule. It also has the largest deviation. But again, there is no obvious reason why the other three analogs (4-F, 4-Br, and 4-NO_2) do not fit to the QSAR.

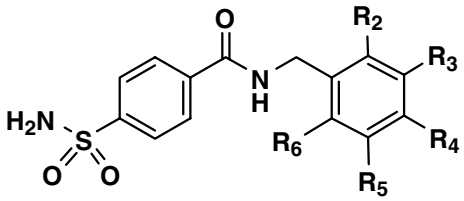
For these examples, and all the other cases in Table 1, the experimentally determined binding modes of the compounds of interest are not known. Therefore, it is not possible to know whether the binding sites and the binding modes of the compounds of interest are the same or at least very similar. Thus, we do not know whether all 15 analogs, including the four outliers in the above examples, bind at the same site of the protein and interact in the same or similar binding manner.

From searching the RCSB protein data bank [3], however, we found a number of examples showing that even a very closely related analog binds in a significantly different manner. Such X-ray structures may provide clues for a possible source of at least some of the outliers observed in QSARs. Ten examples are presented below.

Example 1: *N*-(4-Aminosulfonyl)-*N*-(phenylmethyl)-benzamide analogs as carbonic anhydrase II inhibitors

The structural type II of this example is similar to the structural type I of Table 2 in that they are both sulfonylamino analogs with a substituted phenyl ring, and that both are carbonic anhydrase inhibitors. Carbonic anhydrase is a zinc-containing enzyme that catalyzes the reversible hydration of carbon dioxide: $CO_2 + H_2O \leftarrow\rightarrow HCO_3^- + H^+$[9]. In a study on the protein ligand affinity in the binding of fluoroaromatic inhibitors to CA-II, Kim et al. [10, 11] assumed that the fluorinated protein ligands bind to the same binding site as a nonfluorinated ligand. This was because the physical dimensions of benzene and fluoro-substituted benzene are very similar. A number of CA-II X-ray

Table 3 Structures of the fluorinated *N*-(4-Aminosulfonyl)-*N*-(phenylmethyl)-benzamide ligands bound to the CA-II and their PDB codes. The numbers in parentheses after the PDB codes are the resolution and R-values for the corresponding X-ray structural determination



(II)

Compd No	R_2	R_3	R_4	R_5	R_6	PDB/color in Fig. 2a, b F131V CA-II mutant	PDB/color in Fig. 2c Native CA-II
3-1	F	H	H	H	H	1G45 (1.83 Å; 0.181) [10]/gray	1G1D (2.04 Å; 0.198) [10]/gray
3-2	F	F	H	H	H	1G46 (1.84 Å; 0.185) [10]/gray	1G52 (1.80 Å; 0.165) [10]/cyan
3-3	F	H	H	H	F	1G48 (1.86 Å; 0.178) [10]/gray	1G53 (1.94 Å; 0.198) [10]/orange
3-4	F	F	F	F	F	1G4J (1.84 Å; 0.183) [10]/gray	1G54 (1.86 Å; 0.182) [10]/green
3-5	H	H	H	H	H	1G4O (1.96 Å; 0.197) [10]/gray	
3-6	H	H	F	H	H	1I9L (1.93 Å; 0.223) [11]/gray, orange	
3-7	F	H	F	H	H	1I9M (1.84 Å; 0.194) [11]/gray, orange	
3-8	F	H	H	F	H	1I9N (1.86 Å; 0.221) [11]/gray	
3-9	F	F	F	H	H	1I9O (1.86 Å; 0.212) [11]/gray	
3-10	F	H	F	H	F	1I9P (1.92 Å; 0.217) [11]/green	
3-11	H	F	F	F	H	1I9Q (1.80 Å; 0.183) [11]/gray	

crystal structures bound with such analogs are now available in the RCSB protein structure data bank. Table 3 lists the structures of the structural type II ligands bound to the CA-II and their PDB codes. The quality of these X-ray structures is good and similar to one another. For an average sized protein, resolution <2.5 Å is reasonable; and R-factor value <0.25 is acceptable, and R-factor value <0.2 is reliable. The overall structure of CA-II is essentially unchanged upon the binding of each inhibitor. The interactions of the benzenesulfonamide group with the binding site residues of CA-II are essentially identical.

Figure 2 is a stereo-pair view of CA-II inhibitors bound to the native CA-II or its F131V mutant. In this and in other figures following, the backbones of the protein structures are first overlapped using the sequence alignment program ClustalW of EMBL-EBI [5], and then for the simplicity reason only the ligands are shown. Figure 2a shows all the compounds in Table 3 (Compounds 3-1 to 3-11) bound to the primary binding site of the F131V CA-II mutant. Except one compound (Compound 3-10, 1I9P, green), all compounds bind to the protein in an essentially identical manner. Even though the binding mode of Compound 3-10 (2,4,6-trifluorophenyl) is expected to be very similar to the other compounds including Compound 3-3 (2,6-difluorophenyl) or 3-4 (2,3,4,5,6-pentafluorophenyl) listed in Table 3, its binding conformation and thus its binding mode is quite different from all the other compounds. The structures were all reported from the same laboratory, and the experimental conditions for the X-ray structural determinations were identical. The crystal space groups (P 2_1, P 1 2_1 1) are the same, and the resolutions are essentially identical. Thus, the unique binding mode of Compound 3-10 did not result from differences in experimental conditions.

Figure 2b shows the binding modes of five compounds bound to the secondary binding site of the CA-II F131V mutant. Two different binding modes can be seen [(Compounds 3-6 and 3-7, orange) and (Compounds 3-4, 3-9, and 3-11, gray)]. Again, the chemical structures of Compounds 3-6 (4-fluoro) and 3-7 (2,4-difluoro) are not particularly different from the others, and thus are not expected to bind differently from the others.

Only four X-ray structures of native CA-II are available with a bound ligand of structural type II. Figure 2c shows the three different binding modes of four analogs (Compounds 3-1, gray; 3-3, orange; and 3-2, cyan and 3-4, green) to the primary binding site of the native CA-II. The binding conformation of fluoro-substituted phenyl ring of Compound 3-3 (2,6-difluoro) is almost 90° from the plane of the corresponding ring of Compound 3-2 (2,3-difluoro) and Compounds 3-4 (2,3,4,5,6-pentafluoro), and 3-1 (2-fluoro) binds in between these two, which is about 45°. Again, the

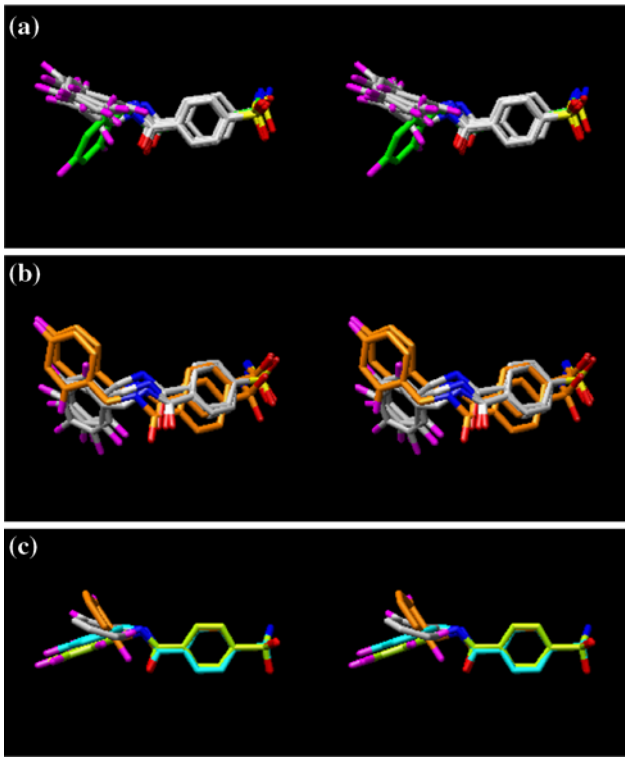


Fig. 2 A stereo-pair view of CA-II inhibitors bound to native or F131V mutant CA-II from the corresponding superimposed inhibitor-bound CA-II structures showing different binding modes. Figure 2a and b show two groups of binding modes for the compounds bound to F131V mutant CA-II, and Fig. 2c shows three different binding modes for the compounds bound to the native CA-II. (**a**) Compounds 3-10 (green) and 3-1 ~ 3-9, 3-11 (gray) bound to the primary binding site of F131V mutant CA-II. (**b**) Compounds 3-6, 3-7, (orange) and 3-4, 3-9, 3-11 (gray) bound to the secondary binding site of F131V mutant CA-II. (**c**) Compounds 3-1 (gray), 3-3 (orange), and 3-2 (cyan), 3-4 (green) bound to the primary binding site of the native CA-II

structures were all reported from the same laboratory, and the experimental conditions for the X-ray structural determinations were identical.

There is no QSAR set in Table 1 that is identical to the present structural set. Therefore, the different binding modes of some of the compounds listed in Table 3 do not provide clear evidence for different binding modes of some of the compounds treated as outliers in the dataset of 6452 and 6453 described above. However, when the binding modes of some simple fluoro-substituted analogs are drastically different from other similar analogs, it is not difficult to imagine that a different binding mode of a compound with a larger substituent than F may be possible.

Compounds with one type of binding mode could give a different SAR or QSAR from the compounds with a different type of binding mode. Different environment in the protein residues surrounding the bound ligands could give different effects on their binding efficacy. Of course, an outlier could arise due to many other reasons. However, the difference in the binding mode of these fluoro-substituted analogs clearly demonstrates a possible source of outliers in QSAR.

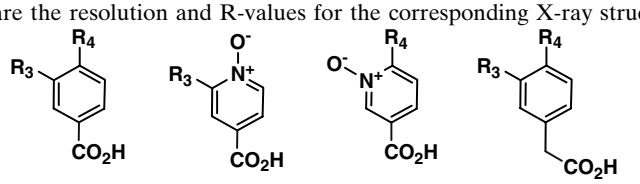
However, several issues need to be considered in this regard. Do the crystal structures represent the real biological binding mode? Is there any uncertainty in the crystal structures? Is the quality similar among different crystal structures? Is there any possibility that the different binding modes of similar compounds are due to the experimental differences in X-ray crystal structure determinations? Some of these issues will be discussed later. Let us first look at some more examples where very close analogs show different binding modes.

Example 2: Hydroxybenzoates and other compounds as protocatechuate 3,4-dioxygenase substrates and inhibitors

Protocatechuate 3,4-dioxygenase (3,4-PCD) catalyzes the oxidative ring cleavage of 3,4-dihydroxybenzoate to produce β-carboxy-cis,cis-muconate [12]. Catecholic dioxygenases are subdivided into two classes based on the site of aromatic ring cleavage and the oxidation state of the catalytically essential metal ion. Extradiol dioxygenases employ a mononuclear Fe^{2+} or Mn^{2+} to catalyze cleavage of the ring adjacent to the vicinal hydroxyl group, and intradiol dioxygenases utilize Fe^{3+} to catalyze ring opening between the hydroxyl groups. 3,4-PCD is an intradiol dioxygenase, and the presence of two vicinal phenolate groups is required for intradiol cleavage of any substrate for all known Fe^{3+}-containing dioxygenases. Crystal structures of *Pseudomonas putida* 3,4-PCD complexed with 11 different ligands currently available in the RCSB protein data bank are listed in Table 4. The quality of these X-ray structures is similar to one another. All the structures were reported from the same laboratory, and the experimental conditions for the X-ray structural determinations were identical.

Figure 3 is a stereo-pair view of ligand-bound protocatechuate 3,4-dioxygenase showing only the bound ligands. Figure 3a shows the first eight compounds in Table 4 (Compounds 4-1 to 4-8). Except one compound (Compound 4-6, 3PCI, green), all the other compounds bind to the protein in an essentially identical manner. Even though the binding modes of Compounds 4-2 (3-fluoro-4-hydroxy, 3PCF, cyan) and 4-5 (p-hydroxyl, 3PCC, cyan) are slightly different from the other compounds, the binding mode of Compound 4-6 (3-iodo-4-hydroxy, 3PCI, green) analog is not expected to be much different from those of the

Table 4 Structures of the various ligands bound to the anaerobic and aerobic 3,4-PCD and their PDB codes. The numbers in parentheses after the PDB codes are the resolution and R-values for the corresponding X-ray structural determination



(III) (IV) (V) (VI)

Compd No	Structural type	R_3	R_4	PDB/color in Fig. 3
4-1	III	OH	OH	3PCA (2.20 Å; 0.165) [12]/gray
4-2	III	F	OH	3PCF (2.15 Å; 0.179) [13]/cyan
4-3	III	OH	H	3PCB (2.19 Å; 0.159) [12]/gray
4-4	III	Cl	OH	3PCH (2.05 Å; 0.167) [13]/ cyan
4-5	III	H	OH	3PCC (1.98 Å; 0.163) [13]/ cyan
4-6	III	I	OH	3PCI (2.21 Å; 0.159) [13]/ green
4-7	IV	OH	–	3PCJ (2.13 Å; 0.173) [12]/orange, 3PCL(2.15 Å; 0.184) [12]/orange
4-8	V	–	OH	3PCK (2.13 Å; 0.173) [12]/orange, 3PCM (2.25 Å; 0.180) [12]/orange
4-9	VI	OH	H	3PCE (2.06 Å; 0.169) [13]/ orange
4-10	VI	H	OH	3PCG (1.96 Å; 0.175) [13]/ green
4-11	VI	OH	OH	3PCN_AA (2.40 Å; 0.166) [14]/cyan, 3PCN_AB (2.40 Å; 0.166) [14]/gray

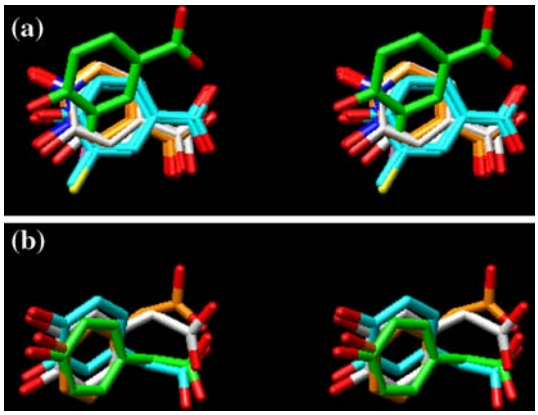


Fig. 3 Stereo-pair views of protocatechuate 3,4-dioxygenase inhibitors from the corresponding superimposed inhibitor-bound protocatechuate crystal structures. (**a**) Binding modes of Compounds 4-1 to 4-8. Except one compound (Compound 4-6, 3PCI, green), all the other compounds bind to the protein in an essentially identical manner. (**b**) Binding modes of Compounds 4-9 (orange), 4-10 (green), and 4-11 (cyan and gray). There are two binding orientations for Compound 4-11

other compounds listed in Table 4. However, the binding orientation of this compound is quite different from all the other compounds. This compound is likely to be an outlier if QSAR is developed from those eight compounds.

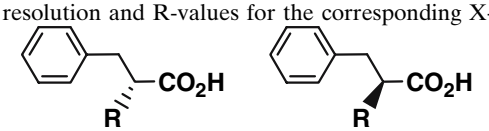
Two compounds in Fig. 3a are not benzoate analogs: one is 2-hydroxyisonicotinic acid *N*-oxide (INO, 3PCJ, 3PCL, orange) and the other is 6-hydroxynicotinic acid *N*-oxide (NNO, 3PCK, 3PCM, orange). These two *N*-oxide inhibitors were designed to mimic the ketonized tautomers of 3,4-dihydroxybenzoate. Crystal structures of these compounds revealed that their binding modes are similar to the first five benzoate compounds (Compounds 4-1 ~ 4-5 in Table 4).

Figure 3b is a stereo-pair view of three substituted phenylacetates: Compounds 4-9 (*m*-hydroxyphenylacetate, 3PCE, orange), 4-10 (p-hydroxyphenylacetate, 3PCG, green), and 4-11 (3,4-dihydroxyphenylacetate, 3PCN_AA, cyan; 3PCN_AB, gray). Two binding modes of Compound 4-11 are reported [14]. Figure 3b shows that there are two distinct groups of binding modes. The two orientations (3PCN_AA and 3PCN_AB) of Compound 4-11 are similar to the binding mode of Compounds 4-10 and 4-9, respectively.

Example 3: Phenylalanine analogs as carboxypeptidase A inhibitors

Carboxypeptidase A (CPA) is a zinc-containing proteolytic enzyme and removes preferentially the C-terminal amino acid residue having an aromatic or hydrophobic aliphatic side chain from peptide substrates [15]. Both D- and L-isomers of *N*-(hydroxyaminocarbonyl)phenylalanine are shown to have strong binding affinity toward CPA [16, 17]. Six crystal structures of CPA complexed with phenylalanine analogs (structural type VII and VIII) available in the RCSB protein data bank are listed in Table 5. The quality of these X-ray structures is not significantly different from one another. Although the structures of Compounds 5-1 and 5-6 were reported by different laboratories from the other compounds, their space

Table 5 Structures of the various phenylalanine analogs bound to the carboxypeptidase A and their PDB codes. The numbers in parentheses after the PDB codes are the resolution and R-values for the corresponding X-ray structural determination



(VII) (VIII)

Compd No	Structural type	R	PDB/color in Fig. 4
5-1	VII	CH_2I	1BAV (1.60 Å; 0.172) [18]/gray
5-2	VII	NHC(=O)NHOH (D)	1HDQ (2.30 Å; 0.139) [17]/green
5-3	VIII	NHC(=O)NHOH (L)	1HEE (1.75 Å; 0.198) [17]/gray
5-4	VIII	$NHSO_2NH_2$	1IY7 (2.00 Å; 0.175) [19]/gray
5-5	VIII	$NHC(=O)NH_2$	1HDU (1.75 Å; 0.198) [17]/orange
5-6	VIII	OH	2CTC (1.40 Å; 0.160) [20]/gray

groups are the same as the other two compounds (Compounds 5-2 and 5-4), and the quality of them is essentially identical.

Figure 4 is a stereo-pair view of the CPA inhibitors from six superimposed inhibitor-bound carboxypeptidase A structures: Compounds 5-1 (1BAV, cyan), 5-2 (1HDQ, green), 5-3 (1HEE, gray), 5-4 (1IY7, gray), 5-5 (1HDU, gray), and 5-6 (2CTC, gray). Figure 4 shows that the binding modes of all phenylalanine analogs listed in Table 5 are essentially identical except Compound 5-2 (1HDQ, green). The iodo group of Compound 5-1 is coordinated to the active site zinc ion and not shown in this figure. Although Compounds 5-1 and 5-2 are stereochemically identical (structural type VII) and the binding position of Compound 5-1 is similar to the other compounds of structural type VIII, Compound 5-2 (1HDQ) binds in a significantly different position from all the others. The binding mode of Compound 5-5 (1HDU, orange) is very similar to the others and identical to that of Compound 5-3 (1HEE).

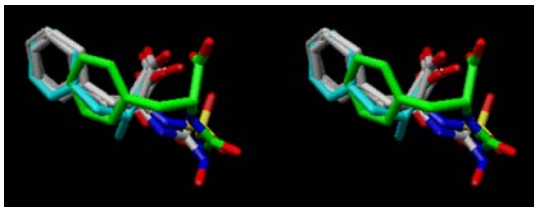
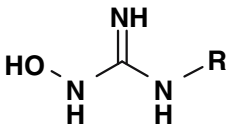


Fig. 4 A stereo-pair view of carboxypeptidase A (CPA) inhibitors from six superimposed inhibitor-bound CPA crystal structures: Compound 5-1 (1BAV, cyan), 5-2 (1HDQ, green), 5-3 (1HEE, gray), 5-4 (1IY7, gray), 5-5 (1HDU, gray), and 5-6 (2CTC, gray). The iodo group of Compound 5-1 is coordinated to the active site zinc ion and not shown in this figure. The binding mode of Compound 5-2 (1HDQ, green) is significantly different from all the others

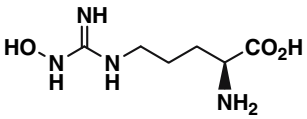
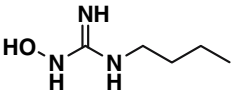
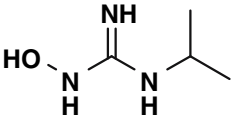
Example 4: *N*-alkyl-*N*′-hydroxyguanidine analogs as neuronal nitric oxide synthase substrates

Nitric oxide synthase (NOS) that catalyzes the formation of nitric oxide from L-arginine exists in three major isoforms: neuronal, endothelial, and immunologic [21, 22] A series of *N*-alkyl-*N*′-hydroxyguanidine compounds (structural type IX) have been characterized as non-amino acid substrates for NOS. The crystal structures of three analogs are available in the RCSB PDB database. Their structures are listed in Table 6. The

Table 6 Structures of the various *N*-alkyl-*N*′-hydroxyguanidine analogs bound to the neuronal nitric oxide synthase and their PDB codes. The numbers in parentheses after the PDB codes are the resolution and R-values for the corresponding X-ray structural determination



(IX)

Compd No	Structure	PDB/color in Fig. 5
6-1		1LZX (2.00 Å; 0.226) [23]/gray
6-2		1M00 (2.05 Å; 0.229) [23]/orange
6-3		1LZZ (2.05 Å; 0.226) [23]/green

structures were all reported from the same laboratory, and the experimental conditions for the X-ray structural determinations were identical. The crystal space groups (P 2_1 2_1 2_1) are the same, and the resolutions and R-values are essentially identical. The novel binding mode of *N*-alkyl-*N*′-hydroxyguanidine to neuronal nitric oxide synthase (nNOS) was described by Li et al. [23]. Crystal structures of the neuronal nitric oxide synthase heme domain complexed with either *N*-isopropyl-*N*′-hydroxyguanidine or *N*-butyl-*N*′-hydroxyguanidine revealed two different binding modes in the substrate binding pocket. The binding mode of *N*-isopropyl-*N*′-hydroxyguanidine was unexpected.

Figure 5 is a stereo-pair view of three hydroxyguanidine nNOS substrates from three superimposed substrate-bound nNOS X-ray crystal structures. While the binding modes of *N*-omega-hydroxy-L-arginine (Compound 6-1, 1LZX, gray) and *N*-butyl-*N*′-hydroxyguanidine (Compound 6-2, 1M00, orange) are very similar, the binding mode of *N*-isopropyl-*N*′-hydroxyguanidine (Compound 6-3, 1LZZ, green) is different from the other two. The position of *N*-hydroxyguanidine group of Compound 6-3 (1LZZ) is flipped 180°. The isopropyl-attached NE atom takes the position normally occupied by the OH-substituted terminal nitrogen NH1 in the other two complexes.

Example 5: Benzyloxycarbonyl analogs as thermolysin inhibitors

Thermolysin (TLN) is a heat-stable extracellular endopeptidase, hydrolyzing the peptide bond on the imino side of the amino acid residue having a large hydrophobic side chain, such as Leu, Ile, and Phe [24]. The enzyme binds one zinc ion at the active site that is required for activity and four calcium ions which are necessary for optimal thermostability [25].

Among the many inhibitor-bound TLN structures, eight crystal structures of TLN complexed with benzyloxycarbonyl analogs (structural types X and XI) are available in the RCSB protein data bank. They are listed in Table 7. The quality of these X-ray structures is excellent and similar one another. The structures were all reported from the same laboratory, and the experimental conditions for the X-ray structural determinations were identical.

Figure 6 is a stereo-pair view of TLN inhibitors from eight superimposed inhibitor-bound TLN structures: Compounds 7-1 (1KR6, gray), 7-2 (1KRO, orange), 7-3 (1KS7, purple), 7-4 (1KTO, green), 7-5 (1KJP, gray), 7-6 (1KKK, gray), 7-7 (1KL6, gray), and 7-8 (1KJO, gray). Figure 6 shows that the binding positions of Compounds 7-5 (1KJP, cyan), 7-6 (1KKK, cyan), 7-7 (1KL6, cyan), and 7-8 (1KJO, cyan) are essentially identical, and those of Compounds 7-1 (1KR6, gray) and 7-3 (1KS7, purple) are also very similar. However, the position of the carbonyl group of Compound 7-2 (1KRO, orange) is different from those mentioned above, and the binding conformation of Compound 7-4 (1KTO, green) is drastically different from all the others. From the binding modes of these compounds, it would not be surprising if Compound 7-4 gives a different SAR or does not fit a QSAR derived from Compounds 7-1, 7-3, 7-5 ~ 7-8 and thus becomes an outlier. Compound 7-2 may also become an outlier.

Example 6: Various ligands of cytochrome c peroxidase

Cytochrome c peroxidase (CCP) is a heme enzyme that catalyzes the oxidation of ferrocytochrome c by peroxides [27]. The W191G cavity of cytochrome c peroxidase is used as a model system for introducing small molecule oxidation in an artificially created cavity. Goodin et al. [28–33] studied the binding modes of several small molecules of structural types XII ~ XVI to the cytochrome c peroxidase. They are listed in Table 8. The quality of these X-ray structures is not much different from one another.

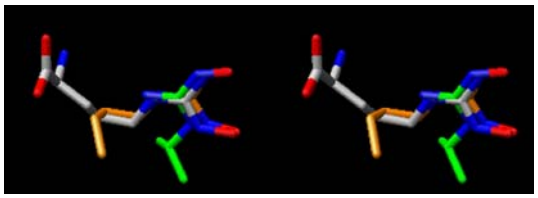


Fig. 5 A stereo-pair view of three hydroxyguanidine nNOS substrates from three superimposed substrate-bound nNOS crystal structures. While the binding modes of Compounds 6-1 (1LZX, gray) and 6-2 (1M00, orange) are very similar, the binding mode of Compound 6-3 (1LZZ, green) is different from the other two by the 180° flipped position of *N*-hydroxyguanidine group

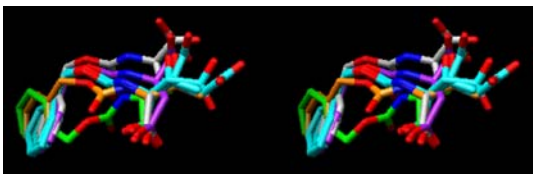
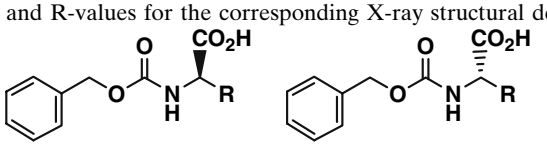


Fig. 6 A stereo-pair view of thermolysin inhibitors from eight superimposed inhibitor-bound thermolysin crystal structures: Compounds 7-1 (1KR6, gray), 7-2 (1KRO, orange), 7-3 (1KS7, purple), 7-4 (1KTO, green), 7-5 (1KJP, cyan), 7-6 (1KKK, cyan), 7-7 (1KL6, cyan) and 7-8 (1KJO, cyan). Binding conformation of Compound 7-4 (1KTO, green) is drastically different from all the other compounds

Table 7 Structures of the benzyloxycarbonyl analogs bound to the thermolysin and their PDB codes. The numbers in parentheses after the PDB codes are the resolution and R-values for the corresponding X-ray structural determination



(X) (XI)

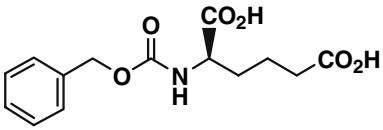
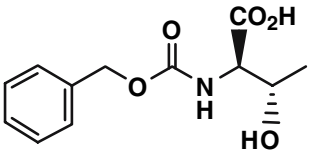
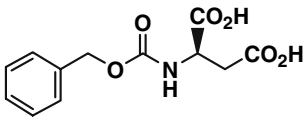
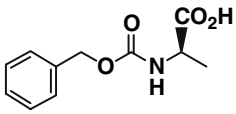
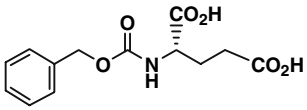
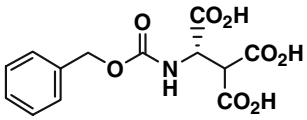
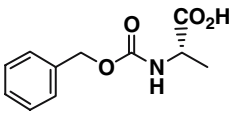
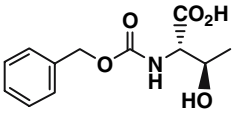
Compd No	Structural type	Structure	PDB/color in Fig. 6
7-1	X		1KR6 (1.80 Å; 0.165) [26]/gray
7-2	X		1KRO (1.70 Å; 0.171) [26]/orange
7-3	X		1KS7 (1.70 Å; 0.170) [26]/purple
7-4	X		1KTO (1.90 Å; 0.167) [26]/green
7-5	XI		1KJP (1.60 Å; 0.176) [26]/cyan
7-6	XI		1KKK (1.60 Å; 0.176) [26]/cyan
7-7	XI		1KL6 (1.80 Å; 0.169) [26]/cyan
7-8	XI		1KJO (1.60 Å; 0.172) [26]/cyan

Figure 7 shows various CCP ligands from the superimposed ligand-bound crystal structures of CCP. Figure 7a is a stereo-pair view of the structural type XII in Table 8. They are Compounds 8-1 (1AC4, gray), 8-2 (1AC8, gray), 8-3 (1AEB, green), and 8-4 (1AED, gray). The binding mode of Compound 8-3 (1AEB, orange) is somewhat different from those of the other three. However, if one focuses one's attention to the orientation of the thiazole ring, the real different one is the binding mode of Compound 8-1 (1AC4, green). The thiazole ring of this compound is rotated clockwise by ~75° from the others.

Figure 7b is a stereo-pair view of the structural type XIII. They are Compounds 8-5 (1AEH, green), 8-6

Table 8 Structures of the various ligands bound to the cytochrome c peroxidase and their PDB codes. The numbers in parentheses after the PDB codes are the resolution and R-values for the corresponding X-ray structural determination

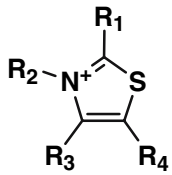
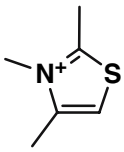
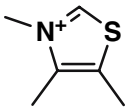
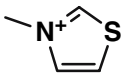
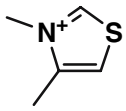
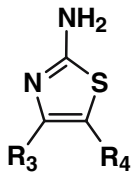
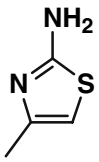
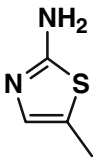
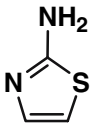
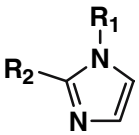
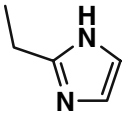
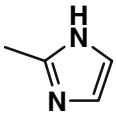
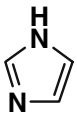
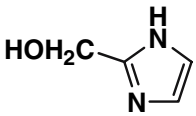
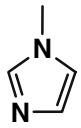
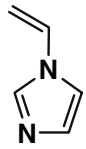
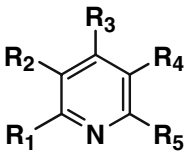
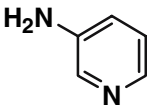
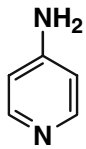
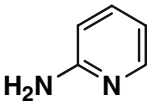
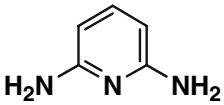
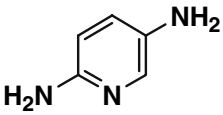
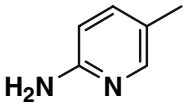
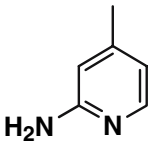
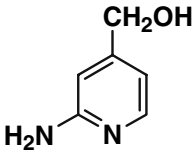
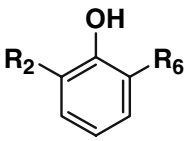
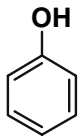
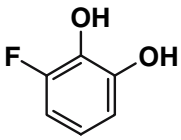
Compd No	Structure	PDB/color in Fig. 7	Compd No	Structure	PDB/color in Fig. 7
					
	(XII)				
8-1		1AC4 (2.10 Å; –) [31]/green	8-2		1AC8 (2.10 Å; –)[31]/gray
8-3		1AEB (2.10 Å; –) [32]/orange	8-4		1AED (2.10 Å; –)[32]/gray
					
	(XIII)				
8-5		1AEH (2.10 Å; –)[32]/green	8-6		1AEN (2.10 Å; –)[32]/gray
8-7		1AEV (2.10 Å; –) [30]/orange			
					
	(XIV)				
8-8		1AEQ (2.10 Å; –)[32]/orange	8-9		1AEU (2.10 Å; –) [29]/orange
8-10		1AES (2.10 Å; –)[29]/green	8-11		2EUU (1.45 Å; 0.146) [33]/orange

Table 8 Continued

Compd No	Structure	PDB/color in Fig. 7	Compd No	Structure	PDB/color in Fig. 7
8-12		1AET (2.10 Å; –)[29]/gray	8-13		1AEJ (2.10 Å; –)[32]/gray
				 **(XV)**	
8-14		1AEF (2.10 Å; –)[32]/orange	8-15		1AEG (2.10 Å; –)[32]/gray
8-16		1AEO (2.10 Å; –)[32]/gray	8-17		2ANZ (1.75 Å; 0.192) [33]/green
8-18		2AQD (1.35 Å; 0.134) [33]/gray	8-19		2EUP (1.40 Å; 0.153) [33]/gray
8-20		2EUT (1.12 Å; 0.143) [33]	8-21		2EUR (1.39 Å; 0.148) [33]
				 **(XVI)**	
8-22		2AS3 (2.40 Å; 0.144) [33]/green	8-23		2AS4 (1.30 Å; 0.138) [33]/gray

(1AEN, gray), and 8-7 (1AEV, orange). It is amazing to see how these three compounds bind to CCP. If one attempted to superimpose these three molecules based on the thiazole ring, one would have never superimposed them as the way shown in Fig. 7b. Their binding modes are completely different from one another even though they are all 2-aminothiazoles, and their binding sites are the same. Not only is the thiazole ring orientation different, but the interaction sites of the amino groups are also completely different. The binding position of N-NH_2 group of Compound 8-5 is significantly different from the other two.

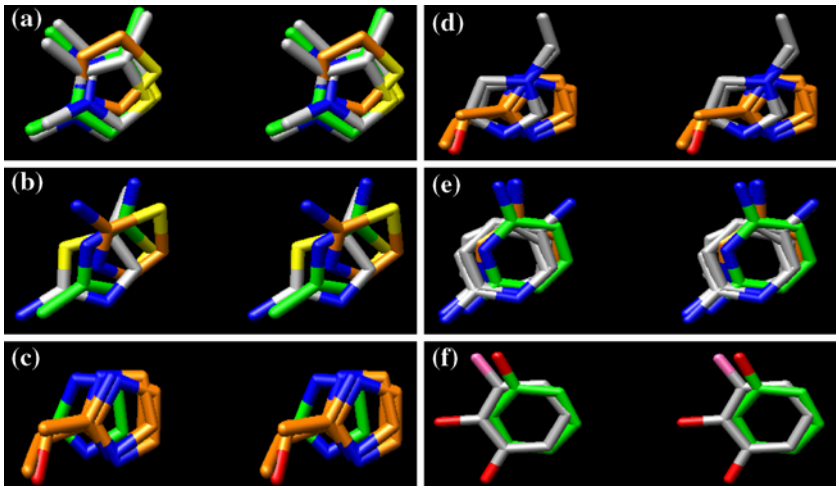


Fig. 7 Stereo-pair views of various ligands from the corresponding superimposed ligand-bound crystal structures of CCP. (**a**) Compounds 8-1 (1AC4, gray), 8-2 (1AC8, gray), 8-3 (1AEB, green), and 8-4 (1AED, gray). The binding mode of Compound 8-3 (1AEB, orange) is somewhat different from those of the other three. (**b**) Compounds 8-5 (1AEH, green), 8-6 (1AEN, gray), and 8-7 (1AEV, orange). Their binding modes are completely different from one another even though they are all 2-aminothiazoles and their binding site is the same. (**c**) Compounds 8-8 (1AEQ, orange), 8-9 (1AEU, orange), 8-10 (1AES, green) and 8-11 (2EUU, orange). The position of Compound 8-9 is somewhat shifted, and its imidazole ring atoms do not overlap with the other three compounds. (**d**) Compounds 8-8 (1AEQ, orange), 8-9 (1AEU, orange), 8-11 (2EUU, orange), 8-12 (1AET, gray), and 8-13 (1AEJ, gray). The binding modes of these compounds show approximately two different binding positions of the imidazole ring. (**e**) Compounds 8-14 (1AEF, orange), 8-15 (1AEG, gray), 8-16 (1AEO, gray), 8-17 (1ANZ, green), 8-18 (2AQD, gray), and 8-19 (2EUP, gray). Pyridine ring nitrogen positions of Compound 8-14 (1AEF, orange) and 8-17 (2ANZ, green) are different each other, and both of them are also different from the other four compounds. (**f**) Compounds 8-22 (2AS3, green) and 8-23 (2AS4, gray). The binding position of the hydroxyl group of phenol is rather close to the position of the 3-fluoro group of 3-fluorocatechol

Figure 7c and 7d are stereo-pair views of the structural type XIV. Figure 7c includes Compounds 8-8 (1AEQ, orange), 8-9 (1AEU, orange), 8-10 (1AES, green) and 8-11 (2EUU, orange). Figure 7c shows that the position of Compound 8-9 is somewhat shifted, and its imidazole ring atoms do not overlap closely with the other three compounds. Figure 7d includes Compounds 8-8 (1AEQ, orange), 8-9 (1AEU, orange), 8-11 (2EUU, orange), 8-12 (1AET, gray), and 8-13 (1AEJ, gray). The binding modes of these compounds show approximately in two different binding positions of the imidazole ring. The figure shows that 2-substituted imidazoles bind one way, and *N*-substituted imidazoles bind another way. Altogether, there are at least three different binding positions observed with the compounds of structural type XIV.

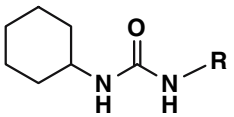
Figure 7e is a stereo-pair view of the structural type XV. They are Compounds 8-14 (1AEF, orange), 8-15 (1AEG, gray), 8-16 (1AEO, gray), 8-17 (1ANZ, green), 8-18 (2AQD, gray), and 8-19 (2EUP, gray). Figure 7e shows that the pyridine ring nitrogen positions of Compounds 8-14 (1AEF, orange) and 8-17 (2ANZ, green) are different each other and both of them are also different from the other four compounds. It shows three groups of the ring nitrogen position. Comparing the NH_2 and pyridine *N* positions of 3-aminopyridine (8-14, 1AEF) and 2,5-diaminopyridine (8-18, 2AQD), one would be tempted to overlap 3-aminopyridyl moiety if one doesn't know the binding modes of these two compounds observed in their crystal structures.

In the studies of Goodin et al. [33] the binding modes of two phenol derivatives (structural type XVI) were also examined: phenol (8-22, 2AS3, green) and 3-fluorocatechol (8-23, 2AS4, gray). Figure 7f shows these two molecules after overlapping the two corresponding protein structures. One may expect that the binding site of the hydroxyl group of phenol may be the same as one of the two hydroxyl group of the 3-fluorocatechol. Surprisingly, however, their crystal structures revealed that this is not the case. The binding position of the hydroxyl group of phenol is rather close to the position of the 3-fluoro group of 3-fluorocatechol.

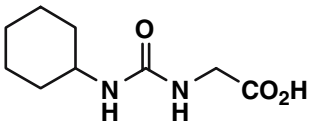
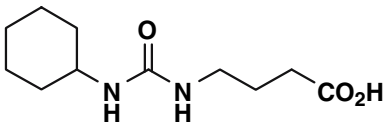
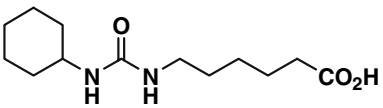
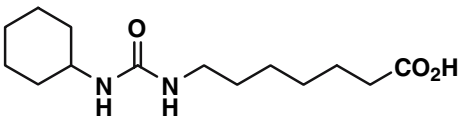
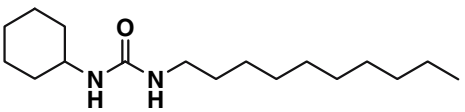
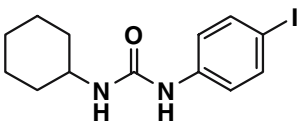
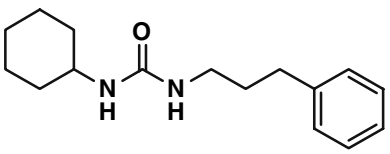
Example 7: Cyclohexylurea analogs as soluble epoxide hydrolase inhibitors

Soluble epoxide hydrolases (sEH) are a class of proteins that catalyze the hydration of chemically reactive

Table 9 Structures of the cyclohexyl urea analogs bound to the soluble epoxide hydrolase and their PDB codes. The numbers in parentheses after the PDB codes are the resolution and R-values for the corresponding X-ray structural determination



(XVII)

Compd No	Structure	PDB/color in Fig. 8
9-1		1ZD2 (3.00 Å; 0.283) [34]/green
9-2		1ZD3 (2.30 Å; 0.212) [34]/gray
9-3		1ZD4 (2.70 Å; 0.233) [34]/gray
9-4		1ZD5 (2.60 Å; 0.231) [34]/orange
9-5		1EK2 (3.00 Å; 0.211) [37]/gray
9-6		1EK1 (3.10 Å; 0.194) [37]/green
9-7		1CR6 (2.80 Å; 0.252) [36]/orange

epoxides to their corresponding dihydrodiol products [32, 34, 35]. Gomez et al. [34] reported the X-ray crystal structures of human soluble epoxide hydrolase complexed with four different dialkylurea inhibitors bearing pendant carboxylate 'tail' of varying length (structural type XVII). Argiriali et al. [36, 37] reported the X-ray crystal structures complexed with each of the three other cyclohexyl urea analogs. Table 9 lists the structures of these inhibitors. The quality of these X-ray structures is similar one another, but overall the quality of all compounds is not so good.

Figure 8a is a stereo-pair view of dialkylurea inhibitors from four superimposed inhibitor-bound human sEH crystal structures: 4-(3-cyclohexylureido)-ethanoic acid (Compound 9-1, 1ZD2), 4-(3-cyclohexylureido)-butyric acid (Compound 9-2, 1ZD3), 4-(3-cyclohexylureido)-hexanoic acid (Compound 9-3, 1ZD4), and 4-(3-cyclohexylureido)-heptanoic acid (Compound 9-4, 1ZD5). Even though all four compounds are 3-cyclohexylureido carboxylic acid analogs, the binding direction of Compounds 9-1 (1ZD2, green), with the shortest side chain, and 9-4 (1ZD5,

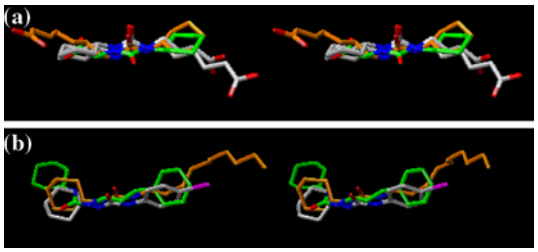


Fig. 8 (**a**) A stereo-pair view of dialkylurea inhibitors from four corresponding superimposed human sEH crystal structures: Compounds 9-1 (1ZD2, green), 9-2 (1ZD3, gray), 9-3 (1ZD4, gray), and 9-4 (1ZD5, orange). Compounds 9-1 and 9-4 bind in opposite direction to that of Compounds 9-2 and 9-3. In addition, the position of the carbonyl group in Compound 9-1 is significantly different from the others. (**b**) A stereo-pair view of three substituted urea inhibitors bound to murine sEH from the corresponding superimposed crystal structures of sEH: Compounds 9-5 (1EK2, orange), 9-6 (1EK1, gray), and 9-7 (1CR6, green). Compound 9-7 is bound to murine sEH in an energetically unfavorable cis-amide linkage

orange), with the longest side chain, is opposite to that of Compounds 9-2 (1ZD3, gray) and 9-3 (1ZD4, gray). Considering the other compounds shown in Fig. 8b discussed below, the binding modes of Compounds 9-1 and 9-4 are unusual. The position of the carbonyl group in Compound 9-1 is also significantly different from those of the others regardless of their binding directions.

Figure 8b is a stereo-pair view of three substituted urea inhibitors, *N*-cyclohexyl-*N*′-decylurea (Compound 9-5, 1EK2, orange), *N*-cyclohexyl-*N*′-(4-iodophenyl)urea (Compound 9-6, 1EK1, gray), and *N*-cyclohexyl-*N*′-propylphenylurea (Compound 9-7, 1CR6, green) bound to murine sEH. Figure 8b shows that the binding mode of Compound 9-7 is drastically different from the other two. Not only is the position of the carbonyl group of this compound different, but the conformation of the amide group is also in an energetically unfavorable cis linkage.

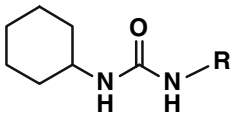
Table 10 shows the data set 11772 and its QSAR results: log $^1/_C$ = 3.29·Clog*P* – 0.80, N = 4, s.d. = 0.079, R^2 = 0.996 (1 outlier), where C is the inhibitory concentration (IC_{50}) of 1-cycloalkyl-3-alkylurea derivatives against mouse epoxide hydrolases. Compound 10-5 in Table 10 and Compound 9-3 in Table 9 are the same. It is interesting to note that the observed outlier behavior of Compound 10-5 is consistent with the unusual binding mode of Compound 9-3. However, the binding mode of Compound 9-3 was determined for human sHE, whereas Compound 10-5 was tested for mouse EH. Therefore, the unusual binding mode of Compound 9-3 may not be firm evidence to explain the outlying behavior of Compound 10-5.

Example 8: Quinazolinediamine analogs as dihydrofolate reductase inhibitors

Dihydrofolate reductase (DHFR) catalyzes the reduction of dihydrofolate to tetrahydrofolate and the conversion of folate to tetrahydrofolate. The reaction is NADPH-dependent [38]. Whitlow et al. [39, 40] studied the binding modes of several quinazolinediamines (structural type XVIII) to the dihydrofolate reductase. They are listed in Table 11. The structures were all reported from the same laboratory, and the experimental conditions for the X-ray structural determinations were identical. The crystal space groups (P 2_1(P 1 2_1 1)) are the same, and the resolutions and R-values are essentially identical.

Figure 9 is a stereo-pair view of five quinazolinediamine analogs from the corresponding superimposed inhibitor-bound DHFR structures. Two compounds with small 4′-substituents [Compounds 11-1 (1IA1,

Table 10 Observed, predicted log 1/C values and parameter table for dataset 11772 [2]



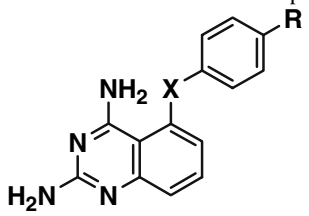
(XVII)

11772: log $^1/_C$ = 3.29·Clog*P* – 0.80, N = 4, s.d. = 0.079, R^2 = 0.996 (outlier = 1)

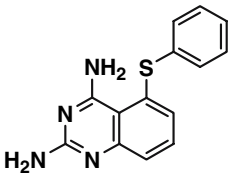
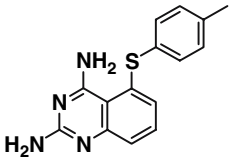
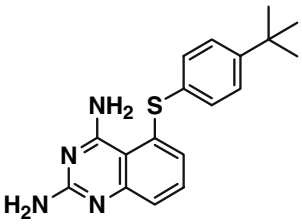
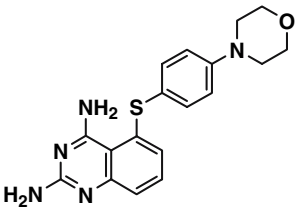
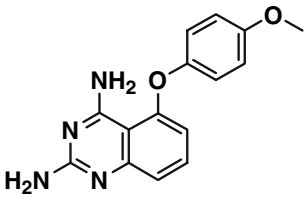
No	Compounds	log $^1/_C$			Clog*P*
		Observed	Predicted	Dev	
10-1	CH_2COOCH_3	4.48	4.49	–0.01	1.60
10-2	CH_2CH_2COOH	3.91	3.95	–0.04	1.44
10-3	$CH_2CH_2COOCH_3$	5.60	5.51	0.09	1.91
10-4	$CH_2CH_2CH_2COOCH_3$	6.48	6.54	–0.04	2.23
10-5*	$(CH_2)_5COOH$	4.05	6.39	–2.34	2.18

* Not included in deriving QSAR

Table 11 Structures of the five quinazolinediamines bound to dihydrofolate reductase and their PDB codes. The numbers in parentheses after the PDB codes are the resolution and R-values for the corresponding X-ray structural determination



(XVIII)

Compd No	Structure	PDB/color in Fig. 9	Compd No	Structure	PDB/color in Fig. 9
11-1		1IA1 (1.70 Å; 0.156) [39]/gray	11-2		1IA2 (1.82 Å; 0.160) [39]/gray
11-3		1IA3 (1.78 Å; 0.160) [39]/green	11-4		1IA4 (1.85 Å; 0.159) [39]/green
11-5		1M79 (1.70 Å; 0.145) [40]/gray			

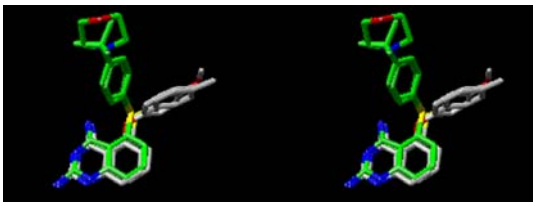


Fig. 9 A stereo-pair view of five quinazolinediamine analogs from the corresponding superimposed inhibitor-bound DHFR structures: Compounds 11-1 (1IA1, gray), 11-2 (1IA2, gray), 11-3 (1IA3, green), 11-4 (1IA4, green), and 11-5 (1M79, gray). Compounds 11-1 and 11-2 bind in one binding mode, whereas Compounds 11-3 and 11-4 bind in another way

gray) and 11-2 (1IA2, gray)] were found to bind with the phenyl group oriented in the plane of the quinazoline ring system. On the other hand, the other two compounds with larger 4′-sustituents [Compounds 11-3 (1IA3, green) and 11-4 (1IA4, green)] were bound to the enzyme with the phenyl group perpendicular to the quinazoline ring as can be seen in Fig. 9. The binding modes of the first two compounds were the one observed previously in other DHFR-inhibitor complexes. However, the binding modes of the later two compounds were found in a manner unprecedented. Compound 11-5 (1M79, gray) also bound to the enzyme similar to Compounds 11-1 and 11-2.

Example 9: Methylpyridazine piperidine alkyloxyphenyl ethyl acetate analogs as poliovirus inhibitors

Picornaviruses, such as the structurally related polioviruses and rhinoviruses, are among the simplest of the animal viruses and are important human pathogens which have been the target of major drug development efforts [41].

Hiremath et al. [42] and Grant et al. [43] studied complexes between capsid-binding, anti-viral drugs methylpyridazine piperidine alkyloxyphenyl ethyacetate analogs (structural type XIX) and poliovirus Type 1 or 3. They are listed in Table 12. The quality of these X-ray structures is similar to one another, but the overall quality is not very good. All the structures were reported from the same laboratory, and the experimental conditions for the X-ray structural determinations were identical. The crystal space groups (P 2_1 2_1 2 for Type 1 and I 2 2 2 for Type 3 poliovirus) are the same as well.

Figure 10 is stereo-pair views of antiviral agents methylpyridazine piperidine alkyloxyphenylethylacetate analogs (structural type XIX) bound to poliovirus Type 1 or Type 3. Compounds 12-1 (1PO2), 12-2 (1VBD), and 12-3 (1PO1) are bound to poliovirus Type 1, and Compounds 12-1 (1VBC), 12-2 (1VBA), and 12-3 (1VBB) are bound to poliovirus Type 3. Figure 10a shows that Compounds 12-2 (VBD, orange) and 12-3 (PO1, gray) overlap by the methylpyridazine piperidine group, not by the phenylethylacetate group. On the other hand, Figure 10b shows that Compounds 12-1 (PO2, gray) and 12-2 (VBD, orange) overlap by the phenylethylacetate group, not by the methylpyridazine piperidine group.

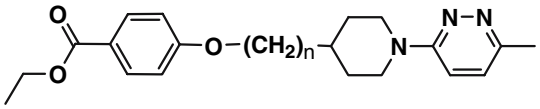
Similar to Fig. 10a for binding to Type 1 poliovirus, Figure 10c shows that Compounds 12-2 (VBA, orange) and 12-3 (VBB, gray) overlap by the methylpyridazine piperidine group, not by the phenylethylacetate group. Similar to Fig. 10b for binding to Type 1 poliovirus, Figure 10d shows Compounds 12-1 (VBC, green) and 12-2 (VBA, orange) overlap by the phenylethylacetate moiety, not by the methylpyridazine piperidine group.

Example 10: Various ligands of structural type XX and XXI as human rhinovirus inhibitors

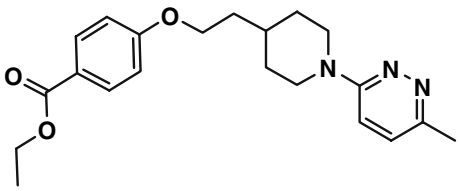
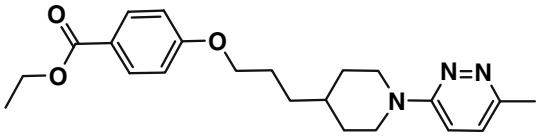
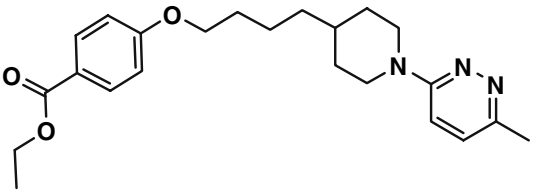
The human rhinovirus (HRV) inhibitors have an icosahedral structures formed by 60 protomers. Each protomer consists of four proteins designated VP1-VP4. The binding site of the capsid binding agents is a hydrophobic pocket formed by VP1 [44].

A number of crystal structures of inhibitor-bound HRV structures are reported [45–50]. Table 13 lists these inhibitors. The quality of these X-ray structures is

Table 12 Structures of the three methylpyridazine piperidine alkyloxyphenyl ethyl acetate analogs bound to poliovirus Type 1 and Type 3 and their PDB codes. The numbers in parentheses after the PDB codes are the resolution and R-values for the corresponding X-ray structural determination



(XIX)

Compd No	Structure	PDB/color in Fig. 10	
		Type 1	Type 3
12-1		1PO2 (2.90 Å; 0.249) [42]/green	1VBC (2.80 Å; 0.289)[43]/green
12-2		1VBD (2.90 Å; 0.212) [43]/ orange	1VBA (2.90 Å; 0.297) [43]/ orange
12-3		1PO1 (2.90 Å; 0.241) [42]/gray	1VBB (2.80 Å; 0.278) [43]/gray

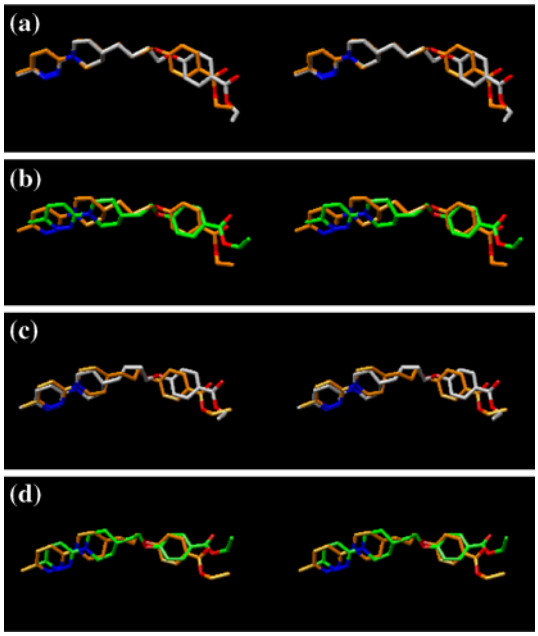


Fig. 10 Stereo-pair views of antiviral agents of methylpyridazine piperidine alkyloxyphenylethylacetate (structural type XIX) bound to poliovirus Type 1 or Type 3: Compounds 12-1 (1PO2), 12-2 (1VBD), and 12-3 (1PO1) bound to poliovirus Type 1, and Compounds 12-1 (1VBC), 12-2 (1VBA), and 12-3 (1VBB) bound to poliovirus Type 3. (**a**) Compounds 12-2 (VBD, orange) and 12-3 (PO1, gray) showing the overlapped position of methylpyridazine piperidine group. (**b**) Compounds 12-1 (PO2, gray) and 12-2 (VBD, orange) showing the overlapped position of phenylethylacetate moiety. (**c**) Compounds 12-2 (VBA, orange) and 12-3 (VBB, gray) showing the overlapped position of methylpyridazine piperidine group. (**d**) Compounds 12-1 (VBC, green) and 12-2 (VBA, orange) showing the overlapped position of phenylethylacetate moiety

similar to one another, but the overall quality is not very good, especially Compounds 13-3 (2HWE) and 13-11 (2HWD). The structures were all reported from the same laboratory, and the experimental conditions for the X-ray structural determinations were identical.

Figure 11 is stereo-pair views of antiviral agents 5-(5-(4-substituted phenoxy)alkyl)-3-methyl isoxazole (structural type XX and XXI) bound to HRV 1A, HRV 14, and HRV 16: Compounds 13-1 through 13-11 bound to HRV 14, Compounds 13-3 and 13-11 to HRV 1A, and Compounds 13-12 through 13-15 to HRV 16.

Figure 11a is a stereo-pair view of Compounds 13-1 (2R06, green), 13-2 (2R07, green), 13-3 (2HWC, green), 13-4 (1R08, green), 13-5 (2R04, green), 13-6 (2RM2, green), 13-7 (2RS5, gray), 13-8 (2RS1. gray), 13-9 (2RR1, gray), 13-10 (2RS3, gray), and 13-11 (2HWD, green). Compound 13-7 (2RS5), 13-8 (2RS1), 13-9 (2RR1), and 13-10 (2RS3) bound in an opposite direction to that of the other eight compounds. Among the eight compounds, the binding position of Compound 13-15 (1NA1) is slightly different from the first seven.

Figure 11b is a stereo-pair view of Compounds 13-1 (1RUC, green) and 13-8 (1RUD, gray) bound to the N1105S mutant of HRV 14. These two compounds are shown to bind in opposite directions.

Figure 11c is a stereo-pair view of Compounds 13-1 (1RUE, and 1RUG) and 13-8 (1RUH and 1RUI), which are bound to a mutant HRV 14. Compound 13-1 (1RUE, green, N1219A; 1RUG, cyan, N1029S) bound in an opposite direction to Compound 13-8 (1RUH, gray, N1219S; 1RUI, orange, S1223G).

Figure 11d is a stereo-pair view of Compounds 13-3 (2HWE, green) and 13-11 (2HWD, gray) bound to HRV 1A. The orientation of the oxazole ring on the right side of these two compounds is different by 180°. However, this result may be doubtful because of their low resolution (3.80 Å). These two compounds are overlapped by the phenyloxazole ring rather than by the methyloxazole at the other end. (Compare with Fig. 11e and f.)

Figure 11e is a stereo-pair view of Compounds 13-12 (1QJU, gray), 13-13 (1QJX, cyan), 13-14 (1QJY, purple), and 13-15 (1ND3, green) bound to HRV 16. There are two groups of positions for the methyl moiety on the 5-membered ring on the right.

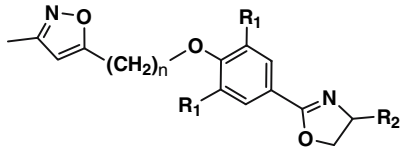
Figure 11f shows three different crystal structures of Compound 13-14 bound to HRV 16. It shows two different binding modes: 1C8M (orange) is one whose coordinates are deposited in PDB but a detailed account is yet to be published by Rossman's lab, and 1ND3 (gray) is one obtained from the crystal obtained by co-crystallization and 1NCR (green) is one obtained from a crystal grown with the ligand present. All three structures were reported by Rossman's lab. It remains to be seen what yielded the two different binding modes for the same compound to the same protein HRV-16.

Figure not presented here is the four different crystal structures of Compound 13-1 (W35; 2R06, 1RUC, 1RUE, 1RUG), which show the same binding mode of this compound to the native and three different HRV 14 mutants. Four different crystal structures of Compound 13-8 (W84; 2RS1, 1RUD, 1RUH, 1RUI) also show the same binding mode of this compound to the native and three different HRV 14 mutants. Likewise, the two structures of Compound 13-15 (W11) bound to HRV 14, 1NA1 from grown_with and 1NCQ from soaking show essentially the same binding mode.

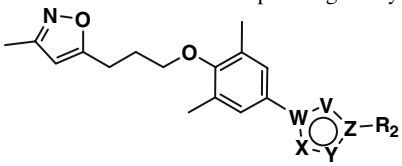
Issues with crystal structures

There are several issues need to be considered with regard to the crystal structures as a possible source of

Table 13 Structures of the various ligands of structural type XX and XXI bound to the rhinovirus 1A, 14, or 16 and their PDB codes. The numbers in parentheses after the PDB codes are the resolution and R-values for the corresponding X-ray structural determination



(XX)



(XXI)

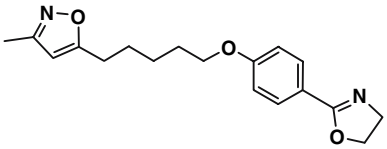
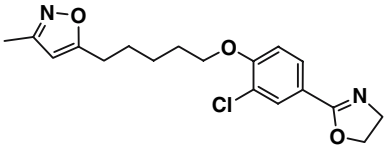
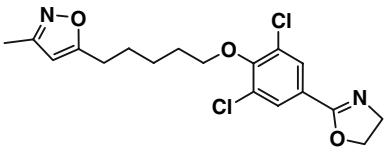
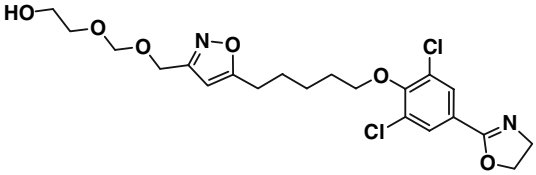
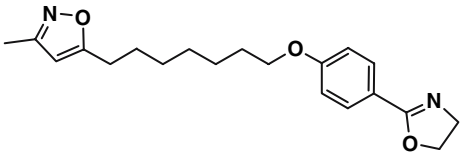
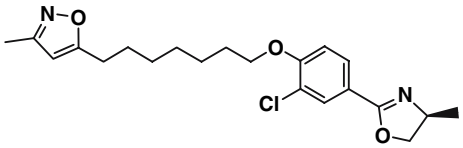
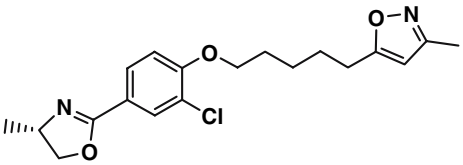
CompdNo	Structure	PDB/color in Fig. 11		
		HRV 1A	HRV 14	HRV 16
13-1 W35			2R06 (3.00 Å; –) [45]/green 1RUC (3.10 Å; –) [49](N1105S)/green 1RUE (2.90 Å; –)[49] (N1219A)/green 1RUG (3.00 Å; –)[49] (N1029S)/cyan	
13-2 W33			2R07 (3.00 Å; –)[45]/green	
13-3 W54		2HWE (3.80 Å; –)[47]/green	2HWC (3.00 Å; –)[47]/green	
13-4 W42			1R08 (3.00 Å; –)[45]/green	
13-5 W71			2R04 (3.00 Å; –)[45]/green	
13-6 W43			2RM2 (3.00 Å; –)[45]/green	
13-7 W56			2RS5 (3.00 Å; –)[45]/gray	

Table 13 continued

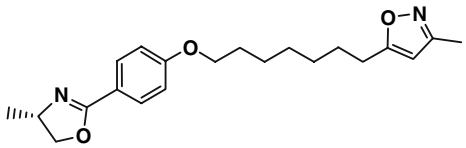
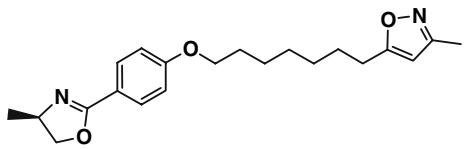
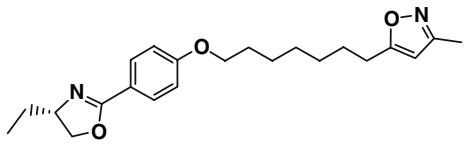
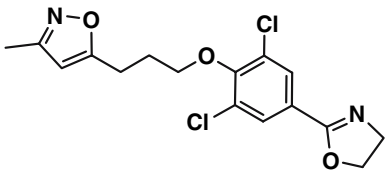
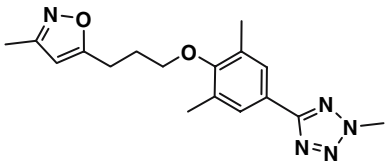
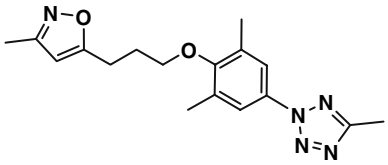
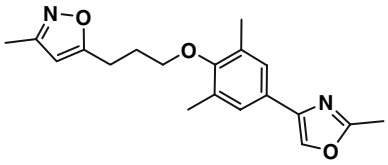
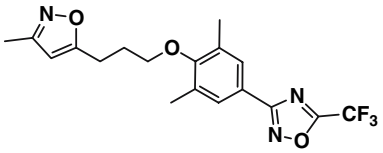
CompdNo	Structure	PDB/color in Fig. 11 HRV 1A	HRV 14	HRV 16
13-8 W84			2RS1 (3.00 Å; –)[45]/gray (S) 1RUD (2.90 Å; –)[49] (N1105S)/gray 1RUH (3.00 Å; –)[49] (N1219S)/gray 1RUI (3.00 Å; –)[49] (S1223G)/orange	
13-9 W8R			2RR1 (3.00 Å; –)[45]/gray (R)	
13-10 W59			2RS3 (3.00 Å; –)[45]/gray (S)	
13-11 W91		2HWD (3.80 Å; –)[47]/gray	2HWB (3.00 Å; –) [47]/green	
13-12 W01				1QJU (2.80 Å; 0.206) [46]/gray
13-13 W02				1QJX (2.80 Å; 0.231) [46]/cyan
13-14(W03)				1QJY (2.80 Å; 0.233) [46]/purple

Table 13 continued

CompdNo	Structure	PDB/color in Fig. 11 HRV 1A	HRV 14	HRV 16
13-15 (W11)			1NA1 (3.30 Å; 0.223) [48]/cyan (grown_with)*	1C8M (2.80 Å; 0.243) [50]/gray
			1NCQ (2.50 Å; 0.217) [48]/(soaking)***	1ND3 (2.80 Å; 0.247) [48]/gray (cocrystal)** 1NCR (2.70 Å; 0.247) [48]/green (grown_with)*

Note that Compounds 13-5 and 13-8 were reported to bind in opposite directions [45], contrary to earlier results [51]

* From crystal grown-with the inhibitor present

** From crystal obtained by co-crystallization method

*** From crystal obtained by soaking method

outliers in SAR or QSAR. Do the crystal structures represent the real biological binding mode? Is there any uncertainty in the crystal structures? Is there any possibility that the different binding modes of similar compounds are due to the differences in the experimental conditions or crystal forms in X-ray crystal determinations? Is the quality similar among different structures? Some of these were considered above, and others will be discussed here.

The crystal structures, or NMR structures in that matter, are the results of a snapshot of the actual binding mode and may not represent, in some cases, the real binding structure [52]. Nevertheless, many structure-based drug design studies treat X-ray crystal structures as actual biological binding structures and use them to optimize lead compounds. The structures of certain proteins [53, 54] determined in crystal by X-ray crystallography and in solution by NMR spectroscopy were found to have identical molecular architectures. On the other hand, those of other proteins [52, 55] indicated some potential differences between the structures in solution and in the crystalline state. (Some suggested, however, that the highest quality NMR structures have accuracies comparable to 2.0–2.5 Å X-ray crystal structures [56, 57]). Quality of some X-ray structures presented in this study, especially those of human rhinovirus inhibitors bound to various HRV, is low: the resolution of the structures of Compounds 13-3 (2HWE) and 13-11 (2HWD) is especially low, 3.80 Å. Also, the R-values of many of the structures are not available. Therefore, some of the observed differences, such as the orientation of the oxazole ring in Fig. 11d, may not be definite. However, the quality of the most other structures is reasonable and consistent, particularly within each set.

Structures obtained from 'soaking' sometimes differ from structures produced by 'co-crystallization' [58, 59]. Even crystals obtained under similar conditions produced minor motions [60]. Nonetheless, the binding sites often remain unchanged. Two structures of Compound 13-15 (W11) bound to HRV 14, 1NA1 from 'grown_with' and 1NCQ from 'soaking' show essentially an identical binding mode. Likewise, 1ND3 (gray) is one obtained from 'co-crystallization' and 1NCR (green) is one obtained from a crystal grown with the ligand present (Fig. 11f). Both essentially have an identical binding mode.

Number of ligand-bound X-ray data sets with different binding modes

Over the last four decades, classical QSAR studies have been conducted based on the assumption that the structurally close analogs bind to the same binding site of the protein in a similar binding mode when ligand-protein interactions are assumed to be involved. The observed biological activity is mostly due to the interaction of the ligand with the protein. Therefore, when the orientation of the functional group of one or more ligands is different, the descriptors for that functional group are not being calculated to reflect the same ligand-protein interactions.

Currently, there are over 40,000 protein structures in the RCSB Protein Data Bank [3]. It is not known how many of them show different binding modes for structurally close ligands. In this study, the binding modes of individual ligands for 14 structural data sets were initially examined as described in the experimental section. The structures of ten of the 14 sets were relatively simple

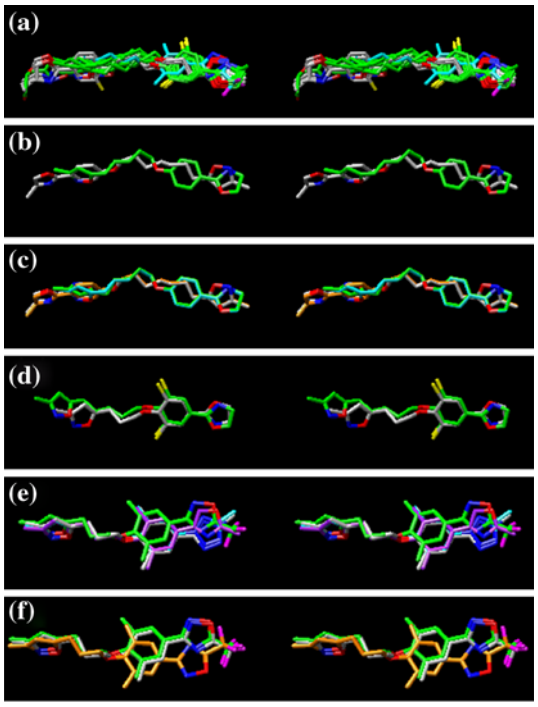


Fig. 11 Stereo-pair views of antiviral agents of 5-(5-(4-substituted phenoxy)alkyl)-3-methyl isoxazole (structural type XX and XXI) bound to HRV 1A, HRV 14, and HRV 16: Compounds 13-1 through 13-11 bind to HRV 14, Compounds 13-3 and 13-11 to HRV 1A, and Compounds 13-12 to 13-15 to HRV 16. (**a**) Compounds 13-1 (2R06, green), 13-2 (2R07, green), 13-3 (2HWC, green), 13-4 (1R08, green), 13-5 (2R04, green), 13-6 (2RM2, green), 13-7 (2RS5, gray), 13-8 (2RS1. gray), 13-9 (2RR1, gray), 13-10 (2RS3, gray), and 13-11 (2HWB, green) bound to HRV 14. (**b**) Compounds 13-1 (1RUC, green) and 13-8 (1RUD, gray) bound to the N1105S mutant of HRV 14. (**c**) Compounds 13-1 (1RUE, and 1RUG) and 13-8 (1RUH and 1RUI) bound to a mutant HRV 14. (**d**) Compounds 13-3 (2HWE, green) and 13-11 (2HWD, gray) bound to HRV 1A. (**e**) Compounds 13-12 (1QJU, gray), 13-13 (1QJX, cyan), 13-14 (1QJY, purple), and 13-15 (1ND3, green) bound to HRV 16. (**f**) Three different crystal structures of Compound 13-14 bound to HRV 16 showing two binding modes. 1C8M (orange) is one whose coordinates are deposited in PDB but a detailed account is yet to be published by Rossman's lab, and 1ND3 (gray) is one obtained from the crystal obtained by co-crystallization and 1NCR (green) is one obtained from a crystal grown with the ligand present

and appropriate to demonstrate the multiple binding modes of the very structurally close ligands. It appears to be more common than exception that structurally close ligands show multiple binding modes. Since this is the first study showing that the different binding modes of individual ligand could affect the results of 2D-QSAR, we decided to present many examples rather than a few to demonstrate that multiple binding modes of structurally close ligands are fairly common rather than exceptions. Even though the binding modes of some ligands are different in these examples, however, the binding sites, at least the position of the main chain, remained essentially identical in each set.

Possible solutions and approaches in QSAR for ligands with different binding modes

Now that we know even structurally closely related ligands may have different binding modes, how can one be sure that all the ligands included in SAR or QSAR have the same binding mode as initially assumed? Of course, the surest way is to determine the binding mode of each ligand by X-ray or neutron diffraction crystallography or by NMR techniques. In fact, many laboratories determine significant number of ligand-bound protein structures for their SAR studies. Nevertheless, it is still not practical in most places to determine structures for every ligand-bound protein. Therefore, the traditional way of doing SAR and/or QSAR will certainly continue. Alternatively, one may examine both the classical and 3D-QSARs and compare the results to see whether they indicate similar SAR. While the classical QSAR studies assume the same binding mode, 3D-QSAR studies can include ligands with different binding modes.

Outliers can be an opportunity in drug discovery research

Not only is one or more outliers observed in many QSARs, most QSAR studies have paid little attention to the outliers and left the outliers without much discussion. Nevertheless, outliers can be very important and may even provide new opportunities or new directions in drug discovery research, especially if the observed biological activity is higher than expected by SAR or QSAR. It may also indicate a different mode of mechanism. Certainly, it would be useful to know the causes of such outliers. It is very likely that multiple binding modes of structurally similar ligands could be a source of outliers in SAR or QSAR.

Conclusions

SAR or QSAR studies without ligand-bound protein structural information first assume that analogous compounds bind to the same binding site in the same or similar binding mode. Studying the QSARs obtained from keyword searches of the C-QSAR database show that a surprisingly large number of compounds were treated as outliers in QSARs. A possible explanation for such outliers was sought from

various ligand-bound protein structures available in the RCBS protein data bank. Ten examples are presented in this report, which show one or more unusual binding modes of some analogous molecules.

Several types of unusual binding modes were observed. Some of them resulted from different torsion angles for part of the molecules. The binding modes of some compounds were different from other compounds due to rotation or translation of their positions from the binding position of the other analogs. Some compounds were even found to bind in an opposite direction. Sometimes, more than one binding mode were even observed.

Such multiple binding modes for some of the analogs presented here do not actually provide indisputable evidence for the outliers observed in QSARs in Table 1. However, there is no doubt that compounds with a different binding mode would not give the same QSAR or fit to the QSAR generated from the compounds with a different binding mode. Unusual binding modes observed in X-ray crystal structures for analogous compounds provide a possibility that it could be a source of outliers observed in SAR or QSAR studies. It may still be worth remembering this in lead optimization process despite the fact that the crystal structures are the results of a snapshot of the actual binding mode and may not represent the real binding structure [52].

Multiple binding modes presented here also remind us that the first assumption in SAR or QSAR that analogous compounds bind to the same binding site in the same or similar binding mode should really be the first assumption. One should soon readjust the assumption based on the observed SAR or QSAR in the continuing process of lead optimization.

Of course, the unusual binding mode may not be the only source of outliers. There could be other reasons. Future studies will deal with other possible sources for outliers in SAR or QSAR.

Acknowledgments The author express sincere gratitude to both Professor Corwin Hansch and Dr. Albert Leo for their generous permission to use their C-QSAR database and the BioByte program Bio-Loom. The author thanks to Drs. Albert Leo and Yvonne Martin for their critical evaluation of the manuscript and valuable suggestions. The author dedicates this paper to the lifetime advisors and friends, Drs. Corwin Hansch, Albert Leo, Yvonne Martin, and Gary Grunewald.

References

1. Kurup A (2003) J Comput Aided Mol Des 17:187–196
2. BioByte 201 W. 4th St., #204, Claremont, CA 91711-4707. clogp@biobyte.com. 909-624-5992
3. Berman HM, Westbrook J, Feng Z, Weissig H, Shindyalov IN, Bourne PE (2000) Nucleic Acids Res 28:235–242
4. CCDC Relibase (version 1.3.2, August 2005), Cambridge Crystallographic Data Center, 12 Union Road, Cambridge CB2 1EZ, United Kingdom
5. Thompson J, Jeanmougin F (2003) ClustalW multiple sequence alignment program (Version 1.83, June 2003)
6. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) J Comput Chem 25:1605–1612
7. Kim KH (1995) In: Hansch C, Fujita T (eds) Classical and three-dimensional QSAR in agrochemistry. American Chemical Society, Washington, DC, vol ACS Symposium Series 606, pp 302–317
8. Hansch C (1995) In: Hansch C, Fujita T (eds) Classical and three-dimensional QSAR in agrochemistry. American Chemical Society: Washington, DC, vol ACS Symposium Series 606, pp 254–262
9. Lindskog S (1997) Pharmacol Ther 74:1–20
10. Kim C-Y, Chang JS, Doyon JB, Baird TT Jr, Fierke CA, Jain A, Christianson DW (2000) J Am Chem Soc 122:12125–12134
11. Kim CY, Chandra PP, Jain A, Christianson DW (2001) J Am Chem Soc 123:9620–9627
12. Orville AM, Lipscomb JD, Ohlendorf DH (1997) Biochemistry 36:10052–10066
13. Orville AM, Elango N, Lipscomb JD, Ohlendorf DH (1997) Biochemistry 36:10039–10051
14. Elgren TE, Orville AM, Kelly KA, Lipscomb JD, Ohlendorf DH, Que Jr L (1997) Biochemistry 36:11504–11513
15. Christianson DW, Lipscomb WD (1989) Acc Chem Res 22:62–69
16. Chung SJ, Kim DH (2001) Bioorg Med Chem 9:185–189
17. Cho JH, Kim DH, Chung SJ, Ha N-C, Oh B-H, Choi KY (2002) Bioorg Med Chem 10:2015–2022
18. Massova I, Martin P, de Mel S, Tanaka Y, Edwards B, Mobashery S (1996) J Am Chem Soc 118:12479–12480
19. Park JD, Kim DH, Woo J-R, Ryu SE (2002) J Med Chem 45:5295–5302
20. Teplyakov A (1993) Acta Cryst D 49:534–540
21. Goldstein IM, Ostwald P, Roth S (1996) Vision Res 36:2979–2994
22. Bredt DS, Snyder SH (1992) Neuron 8:3–11
23. Li H, Shimizu H, Flinspach ML, Jamal J, Yang W, Xian M, Cai T, Wen EZ, Jia Q, Wang PG, Poulos TL (2002) Biochemistry 41:13868–13875
24. Reddy AV (1991) Indian J Biochem Biophys 28:10–15
25. Kester WR (1977) Biochemistry 16:2506–2516
26. Senda M, Senda T, Ogi T, Kidokoro S (2002) Acta Cryst A58:C278
27. Skulachev VP (1998) FEBS Lett 423:275–280
28. Fitzgerald MM (1994) Biochemistry 33:3807–3818
29. Fitzgerald MM (1996) Nat Struct Biol 3:626–631
30. Musah RA (1997) Biochemistry 36:11665–11674
31. Musah RA (1997) J Am Chem Soc. 119:9083–9084
32. Musah RA (2002) J Mol Biol 315:845–857
33. Brenk R (2006) J Mol Biol 357:1449–1470
34. Gomez A (2006) Protein Sci 15:58–64
35. Gomez GA (2004) Biochemistry 43:4716–4723
36. Argiriadi MA (1999) Proc Natl Acad Sci USA 96:10637–10642
37. Argiriadi MA (2000) J Biol Chem 275:15265–15270
38. Schweitzer BI, Dicker AP, Bertino JR (1990) FABES J 4:2441–2452
39. Whitlow M, Howard AJ, Stewart D, Hardman KD, Chan JH, Baccanari DP, Tansik RL, Hong JS, Kuyper LF (2001) J Med Chem 44:2928–2932

40. Whitlow M, Howard AJ, Stewart D, Hardman KD, Kuyper LF, Baccanari DP, Fling ME, Tansik RL (1997) J Biol Chem 272:30289–30298
41. Patick AK, Potts KE (1998) Clin Microbiol Rev 11:614–627
42. Hiremath CN, Filman DJ, Grant RA, Hogle JM (1997) Acta Crystallogr D Biol Crystallogr 53:558–570
43. Grant RA, Hiremath CN, Filman DJ, Syed R, Andries K, Hogle JM (1994) Curr Biol 4:784–797
44. Turner RB (2001) Antiviral Res 49:1–14
45. Badger J, Minor I, Oliveira MA, Smith TJ, Rossmann MG (1989) Proteins: Struct Funct Genet 6:1–19
46. Hadfield AT, Diana GD, Rossmann MG (1999) Proc Natl Acad Sci USA 96:14730–14735
47. Kim KH, Willingmann P, Gong ZX, Kremer MJ, Chapman MS, Minor I, Oliveira MA, Rossmann MG, Andries K, Diana GD, Dutko FJ, McKinlay MA, Pevear DC (1993) J Mol Biol 230:206–227
48. Zhang Y, Simpson AA, Ledford RM, Bator CM, Chakravarty S, Skochko GA, Demenczuk TM, Watanyar A, Pevear DC, Rossmann MG (2004) J Virol 78:11061–11069
49. Hadfield AT, Oliveira MA, Kim KH, Minor I, Kremer MJ, Heinz BA, Shepard D, Pevear DC, Rueckert RR, Rossmann MG (1995) J Mol Biol 253:61–73
50. Chakravarty S, Bator CM, Pevear DC, Diana GD, Rossmann MG The refined structure of a piconavirus inhibitor currently in clinical trials, when complexed with human rhinovirus 16. to be published
51. Smith TJ, Kremer MJ, Luo M, Vriend G, Arnold E, Kamer G, Rossmann MG, McKinlay MA, Diana GD, Otto MJ (1986) Science 233:1286–1293
52. Baldwin ET, Weber IT, Charles RS, Xuan J.-C, Appella E, Yamada M, Matsushima K, Edwards BFP, Clore GM, Gronenborn AM, Wlodawer A (1991) Proc Natl Acad Sci USA 88:502–506
53. Braun W, Vasak M, Robbins AH, Stout CD, Wagner G, Kagi JHR, Wuthrich K (1992) Proc Natl Acad Sci USA 89:10124–10128
54. Nilges M, Macias MJ, O'Donoghue SI, Oschkinat H (1997) J Mol Biol 269:408–422
55. Holak TA, Bode W, Huber J, Otlewski J, Wilusz T (1989) J Mol Biol 210:649–654
56. Montelione GT, Zheng D, Huang YJ, Gunsalus KC, Szyperski T (2000) Nat Struct Biol 7:982–985
57. Billeter M (1992) Q Rev Biophys 25:325–377
58. Hiller N (2005) Protein Sci 15:281–289
59. Segelke BW, Forstner M, Knapp M, Trakhanov SD, Parkin S, Newhouse YM, Bellamy HD, Weisgraber KH, Rupp B (2000) Protein Sci 9:886–897
60. Longhi S, Nicolas A, Creveld L, Egmond M, Verrips CT, de Vlieg J, Martinez C, Cambillau C (1996) Proteins: Struct Funct Genet 26:442–458