Journal of Computer-Aided Molecular Design, **16:** 727–740, 2002.
KLUWER/ESCOM

Combining NMR spectral and structural data to form models of polychlorinated dibenzodioxins, dibenzofurans, and biphenyls binding to the AhR

Richard D. Beger*, Dan A. Buzatu & Jon G. Wilkes
Division of Chemistry, National Center for Toxicological Research, Food and Drug Administration, Jefferson, AR 72079, USA

Received 25 June 2002; accepted in final form 6 November 2002

Key words: 3D-QSDAR; aryl hydrocarbon receptor (AhR); ^{13}C NMR; CoSA; CoSASA; CoSCoSA; dioxin; furan; PCB

Abstract

A three-dimensional quantitative spectrometric data-activity relationship (3D-QSDAR) modeling technique which uses NMR spectral and structural information that is combined in a 3D-connectivity matrix has been developed. A 3D-connectivity matrix was built by displaying all possible assigned carbon NMR chemical shifts, carbon-to-carbon connections, and distances between the carbons. Two-dimensional ^{13}C-^{13}C COSY and 2D slices from the distance dimension of the 3D-connectivity matrix were used to produce a relationship among the 2D spectral patterns for polychlorinated dibenzofurans, dibenzodioxins, and biphenyls (PCDFs, PCDDs, and PCBs respectively) binding to the aryl hydrocarbon receptor (AhR). We refer to this technique as comparative structural connectivity spectral analysis (CoSCoSA) modeling. All CoSCoSA models were developed using forward multiple linear regression analysis of the predicted ^{13}C NMR structure-connectivity spectral bins. A CoSCoSA model for 26 PCDFs had an explained variance (r^2) of 0.93 and an average leave-four-out cross-validated variance (q^2_4) of 0.89. A CoSCoSA model for 14 PCDDs produced an r^2 of 0.90 and an average leave-two-out cross-validated variance (q^2_2) of 0.79. One CoSCoSA model for 12 PCBs gave an r^2 of 0.91 and an average q^2_2 of 0.80. Another CoSCoSA model for all 52 compounds had an r^2 of 0.85 and an average q^2_4 of 0.52. Major benefits of CoSCoSA modeling include ease of development since the technique does not use molecular docking routines.

Abbreviations: Three-dimensional quantitative spectrometric data-activity relationship (3D-QSDAR); Three-dimensional quantitative structure data-activity relationship (3D-QSAR); Aryl hydrocarbon receptor (AhR); Comparative molecular moment analysis (CoMMA); Comparative spectral analysis (CoSA); Comparative structurally assigned spectral analysis (CoSASA); Comparative spectral connectivity spectral analysis (CoSCoSA); heteronuclear multiple quantum correlation (HMQC); multiple linear regression (MLR); Nuclear Magnetic Resonance (NMR); nuclear Overhauser effect spectroscopy (NOESY); Polychlorinated biphenyls (PCBs); Polychlorinated dibenzodioxins (PCDDs); Polychlorinated dibenzofurans (PCDFs), principal component multiple linear regression (PC-MLR).

Introduction

Polychlorinated dibenzo-p-dioxins, dibenzofurans, and biphenyls (PCDDs, PCDFs, and PCBs respectively) are organochlorine industrial compounds and byproducts of those compounds that are widely distributed environmental contaminants. PCDDs and PCDFs are byproducts of the incineration of PCBs

*To whom correspondence should be addressed: RB Beger, Division of Chemistry, National Center for Toxicological Research, Food and Drug Administration, Jefferson, AR 72079, USA (Phone: (870) 543-7080; Fax: (870) 543-7686; E-mail: rbeger@nctr.fda.gov)

and other chlorine-containing organic compounds and have never been manufactured intentionally. The majority of PCDF and PCDD compounds are produced from man-made sources [1,2]. Unlike PCDDs and PCDFs, PCBs were manufactured and used primarily as electrical transformer and capacitor coolants or hydraulic and diffusion pump oils [3].

All these compounds have relatively high melting points and low vapor pressures. This makes them chemically stable and non-reactive in the environment, especially when present in soil or other organic sediments. Because of their chemical inertness, they were regarded for years as reasonably safe to use for industrial purposes. However, eventually they were discovered to have considerable biological activity. They have been shown to be toxicants, having a common receptor-mediated mechanism of action [4,5]. Many of these polychlorinated aromatic compounds cause toxic effects after binding to an intracellular cytosolic receptor called the aryl hydrocarbon receptor (AhR) [6,7]. Thymic atrophy, weight loss, immunotoxicity, acute lethality, and induction of cytochrome P4501A1 have all been correlated with the binding of PCDDs, PCDFs, and PCBs to the AhR [8,9]. Therefore, an important step in predicting the toxicity of PCDDs, PCDFs, and PCBs is being able to estimate their binding affinities to the AhR.

Quantitative spectrometric data-activity relationships (QSDAR) are based on the spectral-activity leg in the triangular structure-spectrum-activity relationship. We have developed QSDAR models for molecules binding to the corticosterone binding globulin [10], the aromatase enzyme [11], and the AhR [12]. Similarly, the binding activities of 45 progestogen steroids to a steroid receptor have been quantitatively modeled with simulated ^{13}C NMR spectra by CoSA [13]. A similar method that combined UV spectra with 31 other physiochemical parameters was used to build models of PCBs and cytochrome P4501A-induced activity [14]. Another similar method used infrared EVA molecular descriptors to formulate models of binding activity [15]

The results of the previous ^{13}C NMR CoSA models were excellent, but we believed they could be further improved [10–12]. The most straightforward way to improve 1D CoSA modeling was to add assigned chemical shift information on a 2D structural template. This modeling method was dubbed comparative structurally assigned spectral analysis (CoSASA) [10,11]. Suprisingly, the CoSASA models were not as robust as the structure-lacking CoSA models. This inferior performance was caused by several factors including (1) unanticipated nonlinear effects, (2) a model variance limitation that is also seen with some quantitative structure-activity relationship (QSAR) modeling, and (3) information that is lost for very long side chains not represented in the template [10,11]. Therefore, CoSASA modeling results of PCDD and PCDF binding to the AhR were not published previously because they were not as good as those produced by the CoSA models [12]. A practical problem with CoSASA modeling was that each model was limited to structurally similar congeners that could be represented in relation to a single template.

The present research initiative avoids some of the foregoing problems by using the ^{13}C NMR spectral data for a test compound and adding the molecule's structural connectivity information into a 3D-connectivity matrix. The most comprehensive form of a 3D-connectivity matrix is built by plotting all possible carbon-to-carbon connections (through-bond and through-space) in relation to their assigned carbon NMR chemical shifts. In this type of matrix representation, the x-axis shows the chemical shifts of carbon i, the y-axis shows the chemical shift of carbon j, and the z-axis the distance between carbons i and j (r_{ij}). Representation of a typical organic compound in this way dramatically increases the information content available to use as a basis for pattern recognition. Each carbon-to-carbon connection in the 3D matrix acts as a constraint on the structure of the molecule. The number of carbon-to-carbon constraints in a 3D-connectivity matrix increases as the square of the number of carbon atoms in the molecule. There are 3N-6 degrees of freedom in a molecule. When the number of structural constraints exceeds the number of degrees of freedom, the information in a 3D-connectivity matrix is said to over-determine the structure of a molecule. For molecules the size of PCDFs, PCDDs, and PCBs, it is possible to reduce the detail in the 3D-connectivity matrix without losing access to the implicit structure-activity characteristics available from this way of describing them. One way to reduce the information is to reduce the third (distance) dimension of the 3D matrix into a set of distance categories or 2D spectral planes. The first 2D plane represents the nearest neighbor through-bond connectivity plane. The other 2D planes are constructed by compressing specific interatomic distances along the z-axis into a few distance categories, one containing all short distance atom-to-atom through-space connections, another for medium- and a third for long-distance atom-to-atom through-

space connections. The particular distance intervals used in a model are not predetermined and can be among the parameters adjusted to optimize a model's performance. Once the ranges have been determined, the compressed connectivity spectra can be defined for any number of compounds. Pattern recognition techniques can then be used to develop the associations between patterns in 2D spectral data and the known biological activity of each compound. The known compounds comprise a training set for the pattern recognition. Once trained, the connectivity spectrum for any test compound can be used by the developed pattern relationships to predict that compound's biological activity.

We have used CoSCoSA modeling to form principal component multiple linear regression (PC-MLR) and artificial neural network models of steroids binding to the corticosterone binding globulin [16] and PC-MLR models of steroids binding to the aromatase enzyme [17] that yielded better correlations than previously published CoSA, CoSASA, or CoMFA models. Another procedure that, like the CoSCoSA methods used here, does not require docking of a molecule onto a grid to produce a model is computationally intensive comparative molecular moment analysis (CoMMA) [18]. CoMMA uses molecular mass and charge distribution moments as descriptors for model development. We advocate CoSCoSA not only for the superior quality but also for its low computational cost, the objectivity with which the connectivity spectra are defined, used and its potential for extension to model structurally diverse data sets.

Standard NMR instrumental techniques include 2D ^{1}H-^{1}H COSY [19] experiments in which connectivity relationships through three bonds are found for nearest neighbor protons and displayed in 2-D as off diagonal cross-peaks. The COSY experiment is similar to 2D ^{13}C-^{13}C through-bond nearest neighbor connectivity spectral patterns. Another standard NMR experiment is the ^{1}H-^{1}H two-dimensional nuclear Overhauser effect spectroscopy (NOESY) [20] in which through-space correlations are found for protons that are less than 5 Å away, again expressed as off-diagonal cross-peaks. The volume of a NOESY cross-peak varies as the inverse of the 6th power of the r_{ij} distance between the protons, and depends as well as on the mixing time of the experiment and the number of different NOE spin diffusion pathways available for dipolar magnetization transfer. Since spin diffusion can cause the NOESY volumes to fluctuate widely for two atoms at the same distance, experimentally determined NOESY volume information is not a good candidate for inferring biological activity from statistical analysis. Although, back-calculated NOESY volumes have been statistically compared with the corresponding experimental NOESY volumes to validate the NMR structure of a duplex DNA [21]. Two-dimensional ^{13}C-^{13}C distance spectra that contain through-space connectivity spectral patterns are conceptually similar to NOESY. Since the relationship is synthesized rather than experimental (as used in this work), in principle the 'spectra' can include distance information for inter-atomic distances greater than 5 Å and can use the number of atoms as a reliable surrogate for peak volume. In this work ^{13}C chemical shifts were used to construct the 3D-connectivity matrix because they contain more explicit structural information than ^{1}H chemical shifts [22,23]. Since the matrix used simulated chemical shifts, the practical limitations of setting up long distance ^{13}C experiments were avoided.

This paper uses ^{13}C NMR spectra combined with structural information in the form of through-bond and selected through-space inter-atomic distances, defined relative to particular reference atoms called structural anchors. It demonstrates that such spectra/distance combinations can be used to produce reliable models of PCDF, PCDD, and PCB compounds binding to the AhR. The work also demonstrates successful combination of spectra/distance matrices for all three compound types into a single model.

Procedures

Table 1 column 3 contains previously reported [4,12] log EC_{50} binding activity data used for training these models. For each compound in Table 1, its ^{13}C NMR spectrum was simulated using the ACD Labs CNMR version 5.0 predictor software [24]. QSDAR CoSCoSA and CoSASA modeling can effectively use predicted ^{13}C NMR spectra. The use of predicted rather than experimentally measured NMR chemical shifts was not necessary for developing the CoSCoSA and CoSASA models, but it saved time, money and in this case it also prevented possible toxic exposures. Predicted ^{13}C NMR data points allow for the spectra to have less dependence on the solvent used. For these compounds, there were no chemical shift peaks outside of 107.0 to 159.0 ppm range. Each compound's spectral pattern was defined using as a surrogate for peak volume the number of atoms having chemical

shifts within segments of this range. The segments are referred to as bins and the number of atoms as the population of a bin. This definition combined with the associated inter-atomic distances provided the spectral component of the CoSCoSA models.

The competitive *in vitro* binding affinities EC_{50} of PCDF, PCDD, and PCB compounds have been determined previously using [^{3}H]-2,3,7,8-tetrachlorodioxin as the radioligand and rodent hepatic cytosol as a source of the AhR [6,9,25–28]. These binding affinities represented the biological activity component of the CoSCoSA models.

Figure 1 shows the flow chart for the CoSCoSA modeling procedure. The 1D ^{13}C NMR spectra were predicted for the compounds in Table 1. The chemical shifts and atom assignment were used from the 1D prediction software and all possible carbon interatomic distances were obtained from ACD 3D Viewer software provided with the ACD ^{13}C NMR prediction software. Since the structures of PCDDs and PCDFs are planar and rigid, the 2D mol file coordinates were usable in the ACD 3D Viewer to determine intramolecular distances between carbon atoms. Similarly, the 2D mol file coordinates of PCB structures were used in the 3D Viewer because the PCB compounds used in this study were considered to have a planar structure. Greater care would be required to determine intramolecular distances for very flexible compounds. 2D planes were built from the combined chemical shift and structural information. One 2D plane represented through-bond COSY spectra. Another represented distances from 'anchoring atoms' to atoms in the middle of the compound. The third, long distance plane served the range from the anchor locations to atoms on the opposite end of the compound. All 2D spectral planes were reduced to a 2.0 ppm resolution in both chemical shift dimensions. These choices resulted in 625 bins. A 2.0×2.0 ppm bin size was used in order to multiply populate as many as possible of the 2D bins. That is, for the training set of compounds it was important that each populated bin be represented by more than one molecule so that binding affinity inferences from the resulting model would represent generalizations based on multiple examples rather than rote 'memorization' of single discrete features. To facilitate generalization, bin dimensions can be increased as long as the inclusion of signals in a single bin does not render as equivalent signals from different parts of the molecule that should be distinguished from each other. Confusion would occur if signals were combined that arose from atomic environments that contribute differently

Table 1. In column two is the compounds used in CoSCoSA models of binding to AhR, column three is the structure's experimental Log EC_{50} and column four is the predicted Log EC_{50} from the 52 compound CoSCOSA model.

#	Compound	Experimental log EC_{50}	52 compound predicted log EC_{50}
1	1-Cl-dibenzofuran	−5.53	−6.53
2	2,8-diCl-dibenzofuran	−6.05	−5.79
3	2,3,7-triCl-dibenzofuran	−8.10	−7.86
4	2,3,8-triCl-dibenzofuran	−7.00	−6.53
5	2,6,7-triCl-dibenzofuran	−7.35	−6.53
6	1,2,3,6-tetraCl-dibenzofuran	−7.46	−7.97
7	1,2,3,7-tetraCl-dibenzofuran	−7.96	−7.38
8	1,2,4,8-tetraCl-dibenzofuran	−6.00	−6.53
9	2,3,4,6-tetraCl-dibenzofuran	−7.46	−7.38
10	2,3,6,8-tetraCldibenzofuran	−7.66	−7.38
11	2,3,7,8-tetraCl-dibenzofuran	−8.60	−8.55
12	1,2,3,7,8-pentaCl-dibenzofuran	−8.12	−8.23
13	1,2,3,7,9-pentaCl-dibenzofuran	−7.40	−7.66
14	1,2,4,7,9-pentaCl-dibenzofuran	−5.70	−5.22
15	1,3,4,7,8-pentaCl-dibenzofuran	−7.70	−7.01
16	2,3,4,7,8-pentaCl-dibenzofuran	−8.82	−8.34
17	1,2,4,6,7,8-hexaCl-dibenzofuran	−6.08	−6.56
18	2,3,4,6,7,8-hexaCl-dibenzofuran	−8.33	−8.45
19	1,2,3,4,7,8-hexaCl-dibenzofuran	−7.64	−7.49
20	1,2,3,6,7,8-hexaCl-dibenzofuran	−7.57	−7.01
21	2,3,4,7,9-pentaCl-dibenzofuran	−7.70	−7.38
22	2,3,4-triCl-dibenzofuran	−5.72	−5.44
23	2,3-diCl-dibenzofuran	−6.33	−6.62
24	2,6-diCl-dibenzofuran	−4.61	−4.27
25	2-Cl-dibenzofuran	−4.55	−5.12
26	4-Cl-dibenzofuran	−4.50	−4.27
27	1-Cl-dibenzodioxin	−5.00	−4.85
28	2,8-diCl-dibenzodioxin	−6.49	−6.53
29	2,3,7-triCl-dibenzodioxin	−8.15	−7.90
30	1,3,7,8-tetraCl-dibenzodioxin	−7.10	−7.21
31	2,3,7,8-tetraCl-dibenzodioxin	−9.00	−9.28
32	1,2,3,4,7-pentaCl-dibenzodioxin	−6.19	−6.65
33	1,2,3,4,7,8-hexaCl-dibenzodioxin	−7.55	−7.21
34	1,2,3,7,8-pentaCl-dibenzodioxin	−8.10	−7.21
35	octaCl-dibenzodioxin	−6.00	−6.53
36	1,2,3,4-tetraCldibenzodioxin	−6.88	−6.53
37	1,2,4,7,8-pentaCl-dibenzodioxin	−6.96	−7.21
38	1,2,4-triCl-dibenzodioxin	−5.88	−6.53
39	2,3,6,7-tetraCl-dibenzodioxin	−7.79	−7.90
40	2,3,6-triCl-dibenzodioxin	−7.66	−7.21
41	2,2′,4,4′,5,5′-hexaCl-biphenyl	−5.10	−4.96
42	2,2′,4,4′-teraCl-biphenyl	−4.89	−4.96
43	2,3,3′,4,4′,5-hexaCl-biphenyl	−6.30	−6.53
44	2,3,3′,4,4′-pentaCl-biphenyl	−6.15	−6.53
45	2,3′,4,4′,5,5′-hexaCl-biphenyl	−5.80	−6.53
46	2,3′,4,4′,5-pentaCl-biphenyl	−6.04	−6.13
47	2,3,4,4′,5-pentaCl-biphenyl	−6.38	−6.13
48	2′,3′4,4′,5-pentaCl-biphenyl	−5.85	−6.53
49	2,3,4,4′-tetraCl-biphenyl	−5.55	−6.53
50	2,3,4,5-tetraCl-biphenyl	−4.85	−4.96
51	3,3′,4,4′,5-pentaCl-biphenyl	−7.92	−7.14
52	3,3′,4,4′-tetraCl-biphenyl	−7.37	−7.76

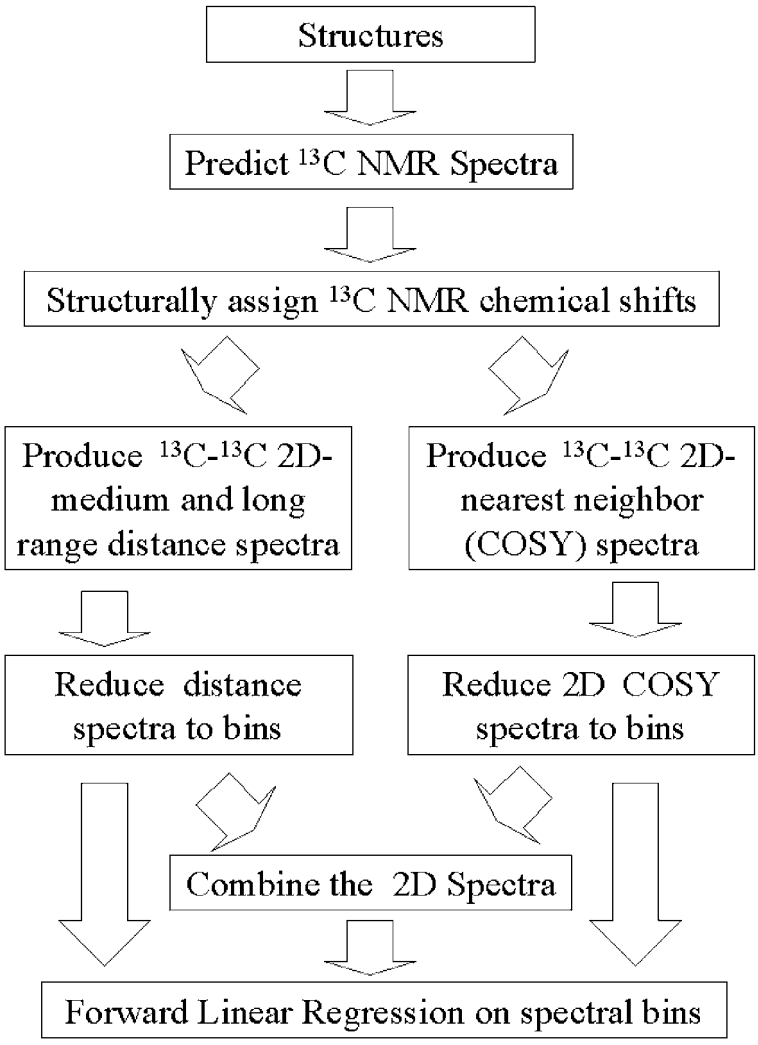

Figure 1. The procedural flow chart for CoSCoSA modeling.

to the biological activity being modeled. Multiply populated bins are required for statistical analysis of the models, particularly model validation. They also reduce the effects of uncertainties from the use of simulated rather than experimental spectra. The first 2D bin included chemical shifts from 157.00 to 158.99 in both the x- and y-dimensions. A 2.0 ppm spectral bin width was chosen in particular because it was used successfully in prior AhR CoSA [12] and CoSCoSA models [16,17]. The raw 2D ^{13}C -^{13}C NMR spectra were represented as two-dimensional bins populated by the number of carbon atoms having a chemical shift within each bin, with the number normalized to a three digit integer. For these rigid compounds each molecule was assumed to have only a single conformation. Typically, a single carbon-to-carbon connectivity on any 2D plane was assigned an area of 100, two carbon-to-carbon connections populating a bin had an area of 200, and so forth.

CoSCoSA models were produced by using the assigned ^{13}C NMR chemical shifts at the 12 carbon positions in the PCDF, PCDD, and PCB molecules, as shown in Figure 2. The blue arrows in Figures 2A, 2B, and 2C show the through-bond COSY carbon-to-carbon connections of the PCDF, PCDD, and PCB molecules, respectively. These through-bond carbon-to-carbon connections were used to simulate 2D ^{13}C-^{13}C COSY spectra of the PCDF, PCDD, and PCB compounds. As mentioned earlier, instead of using all inter-atomic distances that might be included in

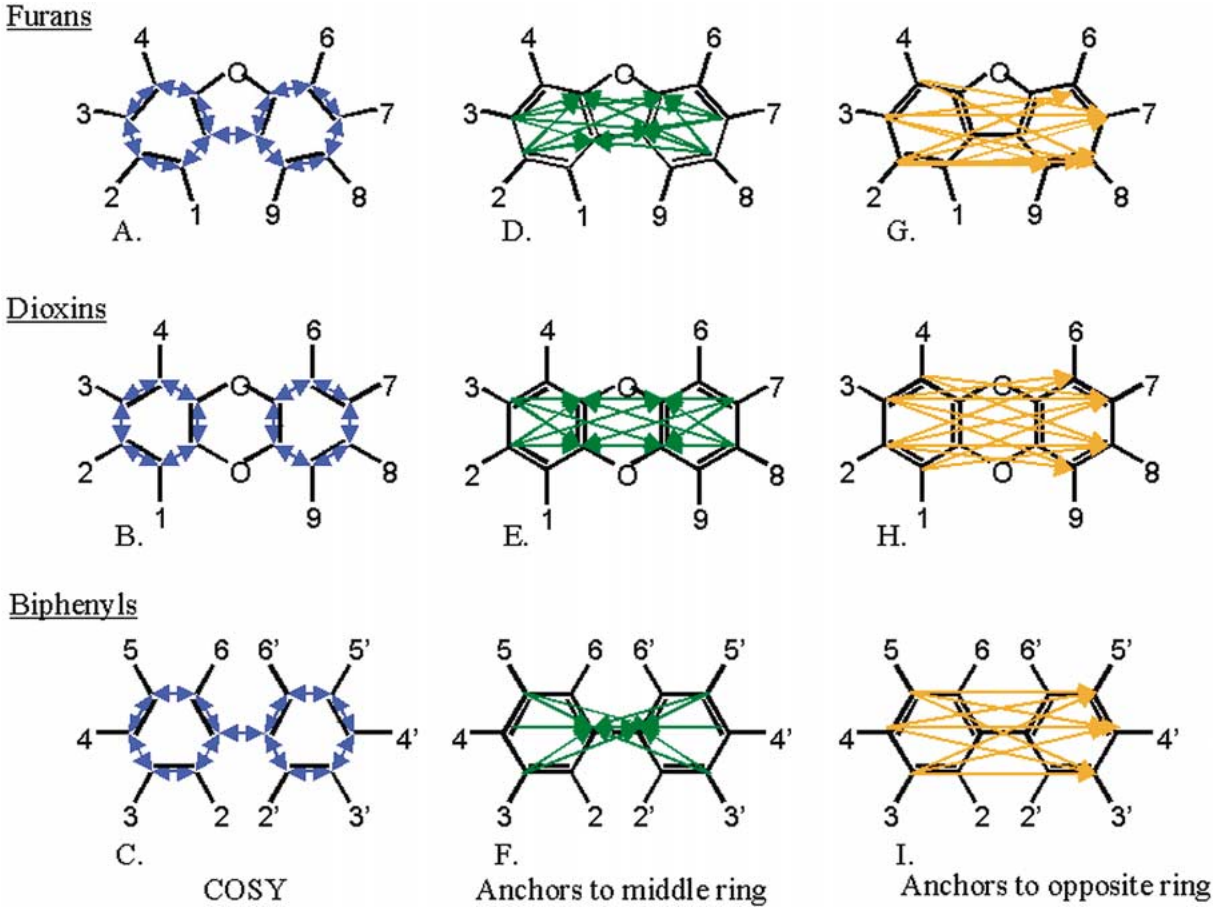

Figure 2. The arrows represent the 2D ^{13}C-^{13}C COSY spectra for PCDFs A. PCDDs B. and PCBs C. The arrows represent middle range 2D ^{13}C-^{13}C distance spectra for PCDFs D. PCDDs E. and PCBs F. The arrows represent long range 2D ^{13}C-^{13}C distance spectra for PCDFs G. PCDDs H. and PCBs I.

the 2D-medium- or long-range planes of the 3D-connectivity matrix, a meaningful subset was defined by breaking each PCDD, PCDF, or PCB molecule into three pieces. The atoms in each compound were segregated into those in an 'anchoring position', those a middle distance from the anchors, and those at a long-range distance – opposite the 'anchor position'.

2,3,7,8-tetrachlorodibenzodioxin is a strong binder in AhR and the presence of these four chlorine atoms constitutes an important factor in determining the compound's toxicity. The distance between positions 2 and 8 or 3 and 7 is 7.0 Å. 2,3,7,8-tetrachlorofuran is also a strong binder to the AhR and the distance between its chlorinated positions 2 and 8 or 3 and 7 is 6.76 Å. Generally, for PCDDs and PCDFs, any pattern that includes chlorination at positions 2, 3, 7 and 8 is known to be associated with the compound's strong binding to the AhR. That is the reason we decided to use the 2, 3, 7, and 8 positions as structural 'anchors' for the distance-related CoSCoSA models.

The green arrows in Figure 2D and 2E show the medium range through-space connections from the 2 and 3 or 7 and 8 anchoring positions to the middle ring carbons in PCDF and PCDD molecules, respectively. PCBs, of course, do not contain a middle ring. For PCB molecules in Figure 2F, we defined anchoring positions as the 3, 4, and 5 positions and the 3′, 4′, and 5′ positions. Atoms in the middle-distance range from these anchors consisted of just the two ring-connecting carbons. The red arrows in Figure 2G and 2H show the long-range through-space connections from the 2 and 3 or 7 and 8 anchoring positions to the opposite ring carbons in PCDF and PCDD molecules, respectively. Likewise, for PCB molecules in Figure 2I, the long-range connections are from the 3, 4, and 5 positions and the 3′, 4′,

and 5′ positions to the carbons on the opposite ring. Since this long-range connectivity interaction overlapped the two anchoring points for each molecule, we choose only one anchor ring as the 'origin' from which all long-range through-space connections originated. CoSCoSA models were built (1) using the nearest neighbor 2D spectral plane only, (2) using the anchoring structural elements through-distance 2D planes, and (3) using a combination of through-bond and through-space information. Since there were two anchors per molecule, we could theoretically have separated the medium-range through-space distance connections from the outer ring anchors to the middle of the compound (D, E, and F) into two separate 2D planes. However, we did not because of symmetry and because the resulting training set would have been too small for meaningful statistics. In contrast, for the short distance 2D COSY, the connectivity arrows point both to and from the nearest neighbor atom. Because of this duality only half of the 2D COSY spectra is needed to define all of the spectra/short distance relationships. Since the medium-range and long-range through-space distances that originate from the anchoring 'origins' do not have the connectivity arrow dual directionality, the whole 2D spectral plane is needed in model development. CoSASA models for PCDD and PCDF compounds were built from 12 assigned carbon chemical shifts as previously described [10,11].

The predicted NMR spectra were calculated by a substructure similarity technique called HOSE [29], which correlates similar 2D substructures with similar NMR chemical shifts. Therefore, errors of estimation produced in the NMR spectra simulation process were propagated through similar structures found in the training set of the QSDAR models. As mentioned above, using the 2 ppm by 2 ppm bin size conveniently reduced the effective error attributable to production of the models based on a HOSE-simulated spectral training set.

All statistical analysis was performed by Statistica version 6.0 software [30]. For each CoSCoSA model, we used forward multiple linear regression (MLR) on selected bins. The models did not use any bins that had less than 2 'hits'. Each CoSCoSA model was built with the goal of comparing performance not only to our previous CoSA models [12] but also to that of the other CoSCoSA models in this paper that used the same set of congeners for training. Therefore, the number of bins selected was not optimized for each particular CoSCoSA model even though for some sets of molecules inclusion of more bins could have increased the F-test value whereas in other models the number used exceeded F-test maximum. For the 26 PCDFs, the following standards were set for the forward MLR. The 'F to Enter'*, a user-defined criterion, was set to 1.0 for the selection of 6 bins. This was done from COSY data alone, for through-space distance from the anchors alone and finally for the through-space distance from the anchors data combined with COSYdata. This produced three corresponding CoSCoSA models. Similarly, for the 14 PCDDs or 12 PCBs the forward MLR was conducted with 'F to enter' set to 1.0 and specifying the selection of 3 bins, CoSCoSA models were produced based on COSY alone, through-space distance alone, and combined spectral data. Finally, a CoSCoSA model of all 52 compounds was produced using forward MLR with 'F to enter' set to 1.0 for the selection of 10 bins from the combined COSY and through-space distance generated 2D spectral planes.

Assessments of these ten CoSCoSA models were achieved using leave-one-out (LOO) or leave-multiple-out cross-validation procedures in which one or more compounds were systematically excluded from the training set and each developed model (missing any contribution from the excluded compound(s)) was used to predict inhibitor binding activities [31]. The cross-validated r^2 (termed q^2) that resulted from the cross-validation experiments was derived from $q^2 = 1 - \text{PRESS/SSD}$. PRESS indicates the sum of the differences between the actual and predicted activity data for each molecule during LOO cross-validation, and SSD is the sum of the squared deviations between the measured and mean activities of each molecule in the training set. During the LOO cross-validation, each compound was removed from the training set and the B-coefficients in the MLR equation were recalculated. This new MLR equation was used to recalculate the log(RBA) of the compound left out. To more rigorously test the validity of the CoSCoSA models, leave-two-out cross validations were performed on the models developed for the 14 PCDDs and the 12 PCBs, and leave-four-out cross-validations were executed on the models for the 26 PCDFs and the multiple compound type model that contained 52 compounds.

*F to enter is a user assigned $F_{critical}$ threshold.

Results and discussion

Tables 2 through 5 report values of n, r^2, q^2, q_n^2, SE (standard error), and F and also identify the 2D bins used for the PCDF, PCDD, PCB, and all 52 compound CoSCoSA models, respectively. For comparison, all of the CoSCoSA models for PCDF compounds were based on 6 MLR-selected 2D bins. All CoSCoSA models for PCDD and PCB compounds were based on 3 objectively selected bins. The CoSCoSA model for all 52 compounds was based on 10 selected bins.

In Table 2 for PCDF compounds, two of three CoSCoSA models had a higher r^2 and q^2 than the corresponding 1D CoSA model using a 2 ppm bin size. Additionally, all three CoSCoSA models for the PCDF compounds had higher r^2 and q^2 than the 2D CoSASA model that associated spectral chemical shift changes at structurally assigned locations with binding to the AhR. The COSY model for PCDF, shown by figure 3A, was based on COSY bins 119–113[†], 125–113, 153–113, 127–119, 155–119, and 127–125. The COSY bin 153–113 identified 10 of the 11 compounds that had activities weaker than −7.0 with only one outlier. The COSY bin 155–119 identified 12 of 15 compounds that had binding activities stronger than −7.0 with 3 outliers. Both the COSY bin at 153–113 and bin 155–119 identified energies of carbon number 5 or 12 from the middle ring connected to the furyl oxygen. These bins are consistent with the fact that bins associated with the furyl oxygen in PCDF compounds also showed a high correlation to binding in our previous 1 ppm and 2 ppm resolution CoSA models [12]. Both the COSY bins at 127–125 and 127–119 identified energies from the carbons numbered 2, 3, 7, and 8. Figure 3D shows performance of the PCDF CoSCoSA model based on medium-range and long-range distance bins. Figure 3G shows the PCDF CoSCoSA model based on COSY, medium-range, and long-range distance bins. The CoSCoSA models for PCDF compounds started with 625 two-dimensional bins in all three 2D planes. When all the bins with only zero population values were removed from the 2D COSY spectral plane, the PCDF CoSCoSA models had 63 remaining bins for that part of the data. Similarly, when all the bins with only zero values were removed from the 2D medium and long range distance planes, the PCDF CoSCoSA models actually had available a combined 133 bins.

[†]All 2ppm bins were written using the format a-b, where a and b are the ppm values corresponding to the two 'connected' atoms.

Our previous 2 ppm resolution CoSA model based on five bins for 26 PCDF compounds had an r^2 of 0.82 and q^2 of 0.72 [12]. A structural parameter model that used Lmax, HOMOs, E(HOMO-LUMO), Log P, and GIW (the geometric analogue of Weiner topological indices) produced by Mekenyan and co-workers [32] was used to produce a 5 component model for 25 PCDF compounds (all 26 PCDF compounds except for 237-trichlorodibenzofuran) with an r^2 of 0.85 and q^2 of 0.71. The best model for 39 dibenzofurans proposed by Turner, one that used three infrared EVA molecular descriptors had an r^2 of 0.96 and a q^2 of 0.73. Another six-component QSAR CoMFA model had an r^2 of 0.85 and a q^2 of 0.72 [15]. These performance results were compared to those of the CoSCoSA PCDF models.

In Table 3 for PCDD compounds, all three CoSCoSA models had r^2 and q^2 values similar to the corresponding 1D CoSA model based on three 2 ppm bins. The CoSCoSA models were based on the selection of 3 2D bins and the previous 1D CoSA model was based on 3 1D bins. All three CoSCoSA models for PCDD compounds had a much higher r^2 and q^2 values than the 2D CoSASA model. Figure 3B shows the COSY model for PCDD compounds that was based on COSY bins 127–123, 141–123, and 143–123. The COSY bin 143–123 correctly identified all 8 compounds with binding activities stronger than −6.96 with only one outlier. The bins at 141 and 143 ppm always identified the energies of one of the four carbon atoms in the middle ring next to the two oxygen atoms. These bins are consistent with the fact that bins in our previous 1ppm and 2 ppm resolution CoSA models associated with the carbon atoms next to the two oxygens in PCDD compounds had a high correlation to binding [12]. Figure 3E shows the PCDD CoSCoSA model based on medium-range and long-range distance bins. While Figure 3H shows the PCDD CoSCoSA model based on COSY, medium-range, and long-range distance bins. The PCDD CoSCoSA models had 48 bins for the ^{13}C-^{13}C COSY data and a combined 54 bins for the medium and long ^{13}C-^{13}C distance connectivity data when all the bins with only zero values were removed from the 2D spectral planes. For comparison, the previous 1D CoSA model based on three chemical shift bins for the 14 PCDD compounds had an r^2 of 0.83 and a q^2 of 0.74 [12]. Five structural parameters were used by Mekenyan and co-workers [31] to produce a five component model for 14 PCDD compounds that had an r^2 of 0.95 and q^2 of 0.82. The model for 25 dibenzodioxins proposed

Table 2. 26 PCDF compound model performance parameters n (parameters used), r^2, q^2, q_4^2, SE, F and MLR Equation (C stands for COSY, M stands for medium-range and L stands for long-range spectra).

Model	n	r^2	q^2	q_4^2	SE	F	MLR equation
1D CoSA [17]	5 Bins	0.82	0.72		0.60	18.6	
2D CoSASA	6 Atoms	0.74	0.70		0.75	9.1	
COSY	6 Bins	0.92	0.84	0.84	0.40	38.7	$-0.03443 * (C_{119-113}) + 0.01245 * (C_{125-113}) + 0.01031 * (C_{153-113}) + 0.00782 * (C_{127-119}) - 0.00426 * (C_{155-119}) + 0.02003 * (C_{127-125})$
Mid + Long	6 Bins	0.83	0.63	0.65	0.61	15.1	$0.00616 * (M_{127-115}) + 0.00576 * (M_{125-117}) - 0.00184 * (L_{119-125}) + 0.00949 * (L_{125-125}) + 0.00967 * (L_{113-127}) + 0.01783 * (L_{121-127})$
COSY + Mid + Long	6 Bins	0.92	0.84	0.84	0.40	38.7	$-0.03443 * (C_{119-113}) + 0.01245 * (C_{125-113}) + 0.01031 * (C_{153-113}) + 0.00782 * (C_{127-119}) - 0.00426 * (C_{155-119}) + 0.02003 * (C_{127-125})$

by Turner that used two infrared EVA molecular descriptors had an r^2 of 0.88 and a q^2 of 0.65, and a two component QSAR CoMFA model had an r^2 of 0.88 and a q^2 of 0.73 [15].

In Table 4 for PCB compounds, all three CoSCoSA models had higher r^2 and much higher q^2 values than the corresponding 1D CoSA model using 2 ppm bins. The CoSA and CoSCoSA models were based on the selection of 3 bins. Figure 3C shows the PCB COSY model based on COSY bins 137–125, 127–127, and 133–131. The COSY bin at 137–125 correctly identified both compounds with a binding activity stronger than −7.0. The COSY bin 137–125 is identified with a bridge carbon having energy associated with 137 ppm and a connecting carbon on one of the rings at 125 ppm. Figure 3F shows the PCB CoSCoSA model based on medium-range and long-range distance bins. Figure 3I shows the PCB CoSCoSA model based on the COSY, medium-range and long-range distance bins. After all the bins with only zero values were removed from the 2D spectral planes, the PCB CoSCoSA models had 28 remaining populated bins available for the ^{13}C-^{13}C COSY data and a combined 39 populated bins for the medium and long ^{13}C-^{13}C distance connectivity data. For comparison, a previous CoSA model based on three bins for the 12 PCB compounds had an r^2 of 0.66 and a q^2 of 0.30 [12]. The model for 33 biphenyls proposed by Turner that used one infrared EVA molecular descriptor had an r^2 of 0.72 and a q^2 of 0.16 and a three component QSAR CoMFA model had an r^2 of 0.87 and a q^2 of 0.49 [14]. Mekenyan and co-workers [32] used structural parameters to produce a model for 12 PCB compounds that had an r^2 of 0.95 and q^2 of 0.79.

In Table 5 results are shown for the single CoSCoSA model of all 52 PCDF, PCDD, and PCBs. Here, the combined COSY and through-space CoSCoSA model had significantly higher q^2 values than the corresponding 1D CoSA model using 2 ppm bins. Additionally, the CoSCoSA model was based only on 10 2D bins, whereas the CoSA models were based on 15 or 12 1D bins, respectively. Figure 4 is a plot of the predicted versus experimental binding for all 52 compounds using the combined COSY, medium and long-range distance spectra. The PCDF compounds are shown with filled circles, the PCDD compounds are shown with open squares, and the PCB compounds are shown with filled triangles. In the combined CoSCoSA model of all 52 compounds, the r^2 for the PCDF compounds is 0.87, the r^2 of the PCDD compounds is 0.84, and the r^2 of the PCB compounds is 0.75. The CoSCoSA model of 52 compounds is composed from 10 2D bins, 3 of which are derived from the COSY spectra, 4 from the medium-range, and the other 3 from the long-range distance spectra. Another interesting feature of this combined model is that each the 10 selected 2D bins contains peaks only found in one of the three compound types. There is a COSY and a medium-range bin that has 'hits' (are occupied) for PCDDs only. There is another COSY and another medium-range bin that has hits for only PCBs. The remaining 6 bins have hits only from PCDFs. Only 6 of the 10 bins (2 bins for each compound type) used in the 52 compound model were used by the previous bin CoSCoSA models for each specific compound type.

Table 6 shows the correlation matrix for the 10 bins used to form the 52 compound CoSCoSA model. In

Table 3. 14 PCDD compound model performance parameters bin n (parameters used), r^2, q^2, q_2^2, SE, F and MLR Equation (C stands for COSY, M stands for medium-range and L stands for long-range spectra).

Model	n	r^2	q^2	q_2^2	SE	F	MLR equation
1D CoSA [17]	3 Bins	0.83	0.74		0.50	16.5	
2D CoSASA	5 Atoms	0.81	0.53		0.60	6.7	
COSY	3 Bins	0.83	0.75	0.74	0.51	15.9	$0.00375 * (C_{127-123}) - 0.01145 * (C_{141-123}) - 0.00794 * (C_{143-123})$
Mid + Long	3 Bins	0.83	0.75	0.71	0.51	16.3	$0.00516 * (M_{123-141}) - 0.00095 * (M_{125-141}) - 0.0017 * (M_{127-141})$
COSY + Mid + Long	3 Bins	0.90	0.79	0.79	0.41	16.2	$-0.0071 * (C_{141-123}) - 0.00527 * (C_{143-123}) + 0.00441 * (M_{123-141})$

Table 4. 12 PCB compound model performance parameters n (parameters used), r^2, q^2, q_2^2, SE, F, and MLR Equation (C stands for COSY, M stands for medium-range and L stands for long-range spectra).

Model	n	r^2	q^2	q_2^2	SE	F	MLR equation
1D CoSA [17]	3 Bins	0.66	0.30		0.63	5.2	
COSY	3 Bins	0.82	0.58	0.58	0.33	12.2	$-0.01714 * (C_{135-125}) + 0.0028 * (C_{127-127}) + 0.00275 * (C_{133-131})$
Mid + Long	3 Bins	0.77	0.66	0.47	0.35	9.1	$0.00275 * (M_{133-135}) - 0.00299 * (L_{133-131}) + 0.00359 * (L_{133-133})$
COSY + Mid + Long	3 Bins	0.91	0.80	0.80	0.28	26.3	$-0.01467 * (C_{137-125}) - 0.00225 * (M_{133-137}) - 0.00301 * (M_{131-139})$

Table 6 there are only two correlations between bins that are greater than 0.25, so there is very little co-linearity between bins used to make the CoSCoSA model. The greatest average correlation between any bin with the other 10 bins was 0.1 and most of the average correlations were much lower than that. The lack of strong correlation among bins suggests that the resulting patterns were based on essentially orthogonal data. The best previous 1D CoSA model for all 52 compounds that was based on the 2 ppm bin size had an r^2 of 0.75 and a q^2 of 0.61 [12]. The current CoSCoSA model had an r^2 of 0.85, a q^2 of 0.73, and a q_4^2 of 0.52 and was based on 10 2D bins selected from the 3D-connectivity matrix. When two outliers are removed from the cross-validation of the 52 compound CoSCoSA model, the q^2 and q_4^2 are 0.77, much improved. Both outliers occurred when a compound had all zeros in every bin except for one column (bin) and that column had only two bin hits in it. When a column with only one remaining 'hit' in it is used during the leave-one-out or leave-four-out cross-validation process, the linear regression B-coefficient can change sign. One needs to be aware of this fact during the development of CoSCoSA models based on selected bins. A six component QSAR CoMFA model of polychlorinated and polybrominated biphenyls, dibenzofurans, and dibenzo-p-dioxins had an r^2 of 0.88 and a q^2 of 0.71 [15]. We conclude that the 10 bin CoSCoSA model for the 52 PCDF, PCDD, and PCB compounds represents an improvement over previously published modeling approaches [5, 12, 15, 32–34].

Almost all the new CoSCoSA models described in this paper (26 PCDFs, 14 PCDDs, 12 PCBs, and combined 52 compounds) produced results at least equivalent to other modeling methods [5,12,15,32–34]. All CoSCoSA models showed some form of improvement over our earlier 1D CoSA models based on the same or fewer 2 ppm bins. We did not publish the results of CoSCoSA models using 1×1 ppm bins, but like the previous CoSA results, generally the r^2 and q^2 improved when going from 2 ppm to 1 ppm bin size. The reason we did not publish the 1 ppm results was that the information was becoming very specific for individual compounds and was not generalized to a set of compounds.

The PCDF and PCDD CoSCoSA models showed major improvement in r^2, q^2, SE, and F-test over

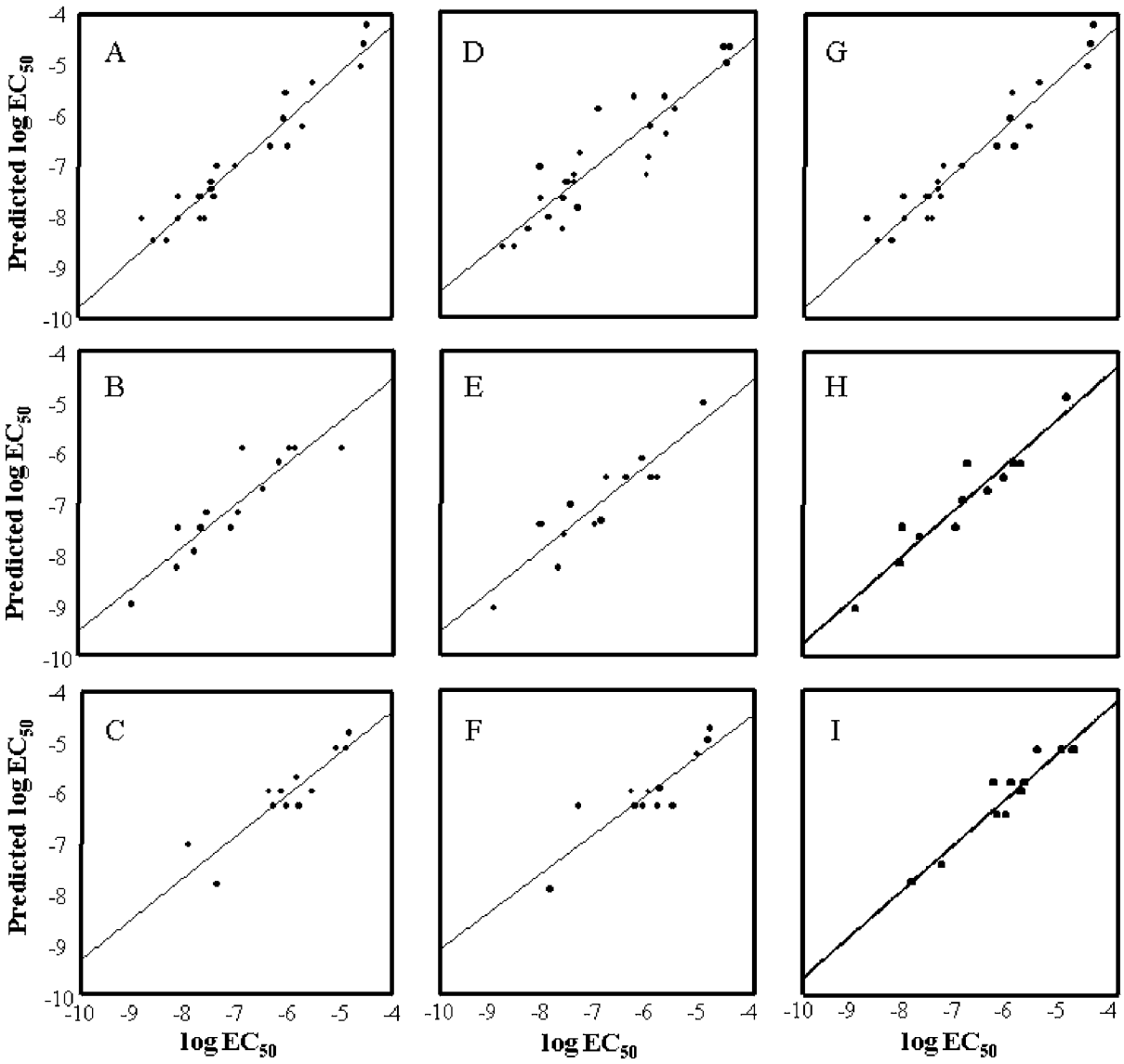

Figure 3. The graphs of the predicted binding versus experimental binding based on COSY spectra for PCDFs A. PCDDs B. and PCBs C. The plots of the predicted binding versus experimental binding based on medium-range and long-range distance spectra for PCDFs D. PCDDs E. and PCBs F. The graphs of the predicted binding versus experimental binding based on the combined COSY and distance spectra for PCDFs G. PCDDs H. and PCBs I.

Table 5. All 52 PCDF, PCDD, and PCB compound model performance parameters n (parameters used), r^2, q^2, q_4^2, SE, F, and MLR Equation (C stands for COSY, M stands for medium-range and L stands for long-range spectra).

Model	n	r^2	q^2	q_4^2	SE	F	MLR equation
1D CoSA [17]	12 Bins	0.75	0.61		0.68	9.9	
COSY + Mid + Long	10 Bins	0.85	0.73	0.52	0.51	24.0	$0.01303 * (C_{121-117}) - 0.00688 * (C_{141-123}) - 0.00617 * (C_{137-125}) + 0.01908 * (M_{119-117}) - 0.00482 * (M_{129-119}) + 0.00392 * (M_{133-135}) + 0.00560 * (M_{123-141}) - 0.01177 * (L_{119-123}) - 0.00852 * (L_{119-125}) + 0.02260 * (L_{121-127})$

Table 6. The correlation matrix for the 10 bins used to form the 52 compound CoSCoSA model.

Bin	1) COSY 121–117	2) COSY 141–123	3) COSY 137–125	4) MED 119–117	5) MED 129–119	6) MED 133–135	7) MED 123–141	8) LONG 119–123	9) LONG 119–125	10) LONG 121–127
1)	1.00	−0.08	−0.04	−0.04	0.05	−0.06	−0.04	−0.05	0.08	−0.07
2)	−0.08	1.00	−0.07	−0.08	−0.15	−0.11	0.00	−0.10	−0.17	−0.12
3)	−0.04	−0.07	1.00	−0.04	−0.07	−0.05	−0.03	−0.05	−0.09	−0.06
4)	−0.04	−0.08	−0.04	1.00	0.05	−0.06	−0.04	0.23	0.42	−0.07
5)	0.05	−0.15	−0.07	0.05	1.00	−0.11	−0.07	−0.10	0.11	−0.13
6)	−0.06	−0.11	−0.05	−0.06	−0.11	1.00	−0.05	−0.08	−0.13	−0.09
7)	−0.04	0.00	−0.03	−0.04	−0.07	−0.05	1.00	−0.05	−0.08	−0.06
8)	−0.05	−0.10	−0.05	0.23	−0.10	−0.08	−0.05	1.00	−0.03	0.46
9)	0.08	−0.17	−0.09	0.42	0.11	−0.13	−0.08	−0.03	1.00	−0.04
10)	−0.07	−0.12	−0.06	−0.07	−0.13	−0.09	−0.06	0.46	−0.04	1.00

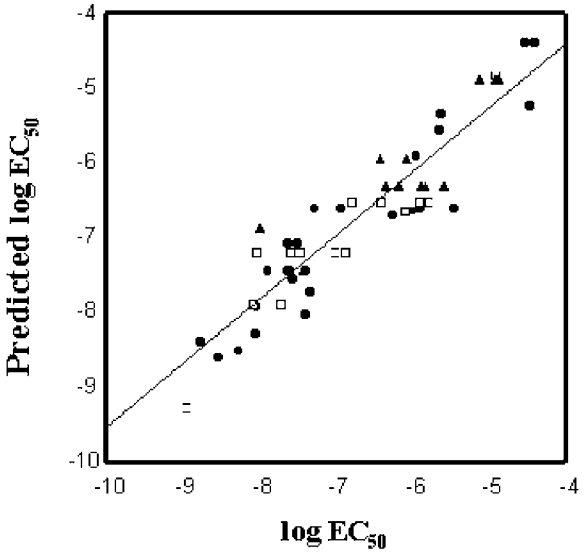

Figure 4. Plot of the predicted binding versus experimental binding for all 52 compounds using the combined COSY plus medium- and long-range spectra. The PCDF compounds are shown with filled circles (●), PCDD compounds shown with open squares (□), and PCB compounds shown with filled triangles (▲).

our own (QSAR-like) CoSASA models, in which the structural information is embedded on a 2D structural template grid. The CoSASA models were based on selected assigned chemical shifts whereas the CoSCoSA models were based on selected 2D Bins from 2D ^{13}C-^{13}C COSY and/or through-space 2D ^{13}C-^{13}C distance spectra. Thus there was more specific structural and spectral information available for the latter group of models. In many of our previous 1D CoSA models of PCDF and PCDD models of binding to AhR, the chemical shifts of carbon positions on the inside middle ring showed some of the strongest correlations to AhR binding. Likewise many of our CoSCoSA models for PCDF and PCDD compounds are based on bins from the middle ring carbons that are connected to the oxygen atom(s). A possible explanation for this is that these bins contain information reflecting the pull on the electron density of the middle ring oxygen atoms by the outer ring chlorine atoms. The number of chlorine substitutions and the position of the substitutions on the outer rings could significantly influence the electron densities of the inner ring carbons through the mechanism described. It is possible that the NMR bins of these carbon atoms are able to record these effects, and that pattern recognition is able to correlate the differences well with the activities of the molecules.

In many of the CoSCoSA models, the overall F score, r^2, and q^2 were still increasing with increasing number of bins used in the model. The continued increase in overall F-test score, q^2 and q^2_n values with bin number argues that over-fitting had not yet occurred and that some continued improvement may be possible by using more bins in the models. In our previous CoSA and CoSCoSA models, we have seen that q^2 begins to level off and drop in value when the F-test score begins to level off or drop off with the addition of more bins or components. To further validate CoSCoSA, more rigorous examination using larger and more diverse test sets will be needed.

From one point of view a 2D ^{13}C-^{13}C COSY spectrum may be considered as a 2D ^{13}C-^{13}C distance spectrum in which the distances between nuclei are less than 1.6 Å. The 2D ^{13}C-^{13}C COSY spectrum

and ^{13}C-^{13}C distance spectra may be viewed as a reduced form of a 3D matrix in which the ^{13}C chemical shifts comprise the x and y axes, and atomic distance lies along the z-axis. For the 2D ^{13}C-^{13}C COSY spectra, all distances less than 1.6 Å are used and projected onto a 2D plane. For the through-space distance spectra in these CoSCoSA models, we obtained information from medium-range and a long-range spectra from positions 2 and 3, or 7 and 8, for PCDF and PCDD compounds, or from positions 3, 4, and 5, or 3′, 4′, and 5′ in PCB compounds. The medium-range and long-range through-space distance information was compressed into separate 2D planes. For this particular modeling task we did not choose to use the structural information under 2.2 Å because the chemical shifts are very dependent on and inherently contain quantum mechanical information from the local structural environment. Much of the success of previous CoSA and CoSCoSA models may be attributed to the fact that our models have only dealt with fairly flat and rigid molecules. A true test of CoSCoSA models will be on a set of non-rigid compounds where the CoSCoSA models can be formed from 4D-connectivity matrices. The 4D-connectivity matrices can be formed from selected sets of snapshots of molecular dynamics trajectories of each compound. Therefore, the through-spatial distance bins in a connectivity matrix for flexible compounds can vary along the z-axis (r_{ij}) from 1 Å to some maximum determined by the outer limits of molecular distension.

The CoSCoSA modeling system can be applied to receptor binding systems for which the structure-activity relationship is unknown, a common situation faced by new drug discovery or lead optimization programs in the pharmaceutical industry. Producing QSAR models without detailed structural binding information is very unreliable and based on docking methods that can effect the reliability of a QSAR model [35]. One difference between CoSCoSA and CoMFA models is that CoMFA is an ‘analog’-like technique that relies on calculating external physical factors, and CoSCoSA modeling is ‘digital’-like and relies on statistics of internal quantum mechanical properties of chemical shifts and atom-to-atom distances. Although the information content the two modeling types are based on is quite similar, the information is manifested in contrasting ways. It’s possible in the future that the CoSCoSA and CoMFA methods could complement each other.

One limitation to CoSCoSA and CoSA modeling is that they need larger training sets to become very predictive. We see that the 26 PCDF compound CoSCoSA models generally had higher F-test scores than the models for 14 PCDD and 12 PCB compounds. Another general observation is the CoSCoSA models with the combined COSY and through-space distance information were more accurate than the models based on COSY or through-space information alone. Because the models for PCBs are based on training sets with only 12 compounds that do not have a lot of continuous diversity in log (EC_{50}), the models appear to be forming small clusters for PCB compounds as seen in Figure 3 (C, F, and I) and Figure 4.

CoSCoSA modeling can be a valuable modeling system for any project that requires predictive structural models, such as new drug lead optimization. Predicted ^{13}C NMR chemical shifts are excellent descriptors of the nearby surrounding environment of an atom in a molecule. By adding nearest neighbor and through-space distance-related structural information to the ^{13}C NMR chemical shifts, we have been able to produce accurate models of binding to AhR. Similarly, we were able to use 2D COSY and through-space long-distance spectral data to form PC-MLR and ANN models of steroids binding to the corticosterone binding globulin [16] and PC-MLR models of steroids binding to the aromatase enzyme [17]. In those papers, the arrows for the through-space distance spectra of the steroids were bi-directional, and the idea of using unidirectional arrows from anchor points was not used. Despite this previous shortcoming, the through-space spectral models still worked for the steroids because there was very little spectral overlap. The combined COSY and long-range distance steroid models yielded better correlations than previously published CoSA, CoSASA, or CoMFA models [16,17]. The idea to use selected molecular structural features as ‘anchors’ in the through-space distance spectra for CoSCoSA can be beneficial to drug design research.

Conclusions

The existence of a ‘hit’ in a particular i, j pair (r_{ij} bin) connotes only that the molecule contains two atoms with chemical shift energies at a particular distance range apart. Thus, the 3D conformational information used in CoSCoSA models is not registered with respect to an assumed structural backbone or assumed number of atoms in the molecule. It follows that, unlike QSAR methods, the 3D-QSDAR models, in either model development or use for prediction, do

not require any assumptions regarding the molecular alignment sequence or other docking events by which a molecule interacts with a biological receptor.

In the work presented here, the size of the two-dimensional bins used from the reduced 3D-connectivity matrix has not been optimized. In any event, without serious optimization efforts we have developed very accurate models of PCDDs, PCDFs, and PCBs binding to the AhR, and steroids binding the aromatase receptor and corticosterone binding globulin [16,17]. In the future, we may be able to use genetic algorithms to optimize the bin sizes used in CoSCoSA models. The CoSCoSA models developed here use predicted ^{13}C NMR spectral data and are not computationally intensive. Furthermore, modifications to the training set, such as the inclusion of other molecules, can be accomplished rapidly. Overall, CoSCoSA modeling is very fast and could be automated for high throughput systems.

References

1. Thomas, V. and Spiro, C., Toxicol. Environ. Chem., 50 (1995) 1.
2. Brzuzy, L.P. and Hites, R.A., Environ. Sci. Technol., 30 (1996) 1797.
3. EPA: Polychlorinated Biphenyls; Criteria Modification; Hearings. Federal Register, 44 (1979) 31514.
4. Safe, S., Crit. Rev. Toxicol., 21 (1990) 50.
5. Mekemyan, O.G., Veith, G.D., Call, D.J. and Ankley, G.T., Environ. Health Perspect., 104 (1996) 1302.
6. Bhandiera, S., Sawyer, T., Romkes, M., Zmudzka, B., Safe, L., Mason, G., Keys, B. and Safe, S., Toxicology, 32 (1984) 131.
7. Mason. G., Farrell, K., Keys, B., Piskorska-Pliszczynska, J., Safe, L. and Safe, S., Toxicology, 41 (1986) 21.
8. Mason. G., Sawyer, T., Keys, B., Bandiera, S., Romkes, M., Piskorska-Pliszczynska, J., Zmudzka, B. and Safe, S., Toxicology, 37 (1985) 1.
9. Bandiera, S., Safe, S. and Okey, A. B., Chem.-Biol. Interact., 39 (1982) 259.
10. Beger, R.D. and Wilkes, J.G., J. Comput.-Aid. Mol. Des., 15 (2001) 659.
11. Beger, R.D. and Wilkes, J.G., J. Chem. Inf. Comput. Sci., 41 (2001) 1360.
12. Beger, R.D. and Wilkes, J.G., J. Chem. Inf. Comput. Sci., 41 (2001) 1322.
13. Andersson, P.L., van der Burght, A.S.A.M., van der Berg, M. and Tysklind, M., Environ. Toxicol. Chem., 19 (2000) 1454.
14. Bursi, R., Dao, T., van Wilk, T., de Gooyer, M., Kellenbach, E. and Verwer, P., J. Chem. Inf. Comput. Sci., 39 (1999) 861.
15. Turner, D.B., Willett, P., Ferguson, A.M. and Heritage, T., J. Comput. Aid. Mol. Des., 11 (1997) 409.
16. Beger, R.D., Buzatu, D.A., Lay, J.O. and Wilkes, J.G., J. Chem. Inf. Comp. Sci. 42 (2002) 1123.
17. Beger, R.D. and Wilkes, J.G., J. Molec. Recognit., 15 (2002) 154.
18. Silverman, B.D. and Platt, D.E., J. Med. Chem., 39 (1996) 2129.
19. Aue, W.P., Bartholdi, E. and Ernst, R.R., J. Chem. Phys., 24 (1976) 2229.
20. Jeener, J., Meier, B.H., Bachmann, P. and Ernst, R.R., J. Chem. Phys., 71 (1979) 4546.
21. Beger, R.D. and Bolton, P.H., Biol. Chem., 273 (1998) 15565.
22. Beger, R.D. and Bolton, P.H., J. Biomol. NMR, 10 (1997) 129.
23. Wishart, D.S. and Sykes, B.D., Meth. Enzymol., 239 (1994) 363.
24. ACD/Labs CNMR software version 5.0, Toronto, Canada.
25. Poland, A. and Knutson, J.C., Annu. Rev. Pharmacol. Toxicol., 22 (1982) 517.
26. Poland, A., Glover, E. and Kende, A.S., J. Biol. Chem., 251 (1976) 493.
27. Safe, S. Crit. Rev. Toxicol., 13 (1984) 319.
28. Safe, S.H. Annu. Rev. Pharmacol. Toxicol., 26, (1986) 371.
29. Bremser, W. HOSE – a Novel substructure Code. Anal. Chim. Acta., 103 (1978) 355.
30. Statisica, StatSoft software, Tulsa, OK.
31. Cramer, R.D., Bunce, J.D. and Patterson, D.E., Quant. Struct.-Act. Relat., 7 (1988) 18.
32. Mekemyan, O.G., Veith, G.D., Call, D.J. and Ankley, G.T. Environ. Health Perspect., 104 (1996) 1302.
33. Rannug, U., Sjogren, M., Rannug, A., Gillner, M., Toftgard, R., Gustafsson, J.-A., Rosenkranz, H. and Klopman, G., Carcinogenesis, 12 (1991) 2007.
34. Kafafi, A.A., Afeefy, H.Y., Said, H.K. and Hakimi, J.M., Chem. Res. Toxicol., 5 (1992) 856.
35. Cho, S.J. and Tropsha A., J. Med. Chem., 38 (1995) 1060.