

Analysis of Pharmacokinetics, Pharmacodynamics, and Pharmacogenomics Data Sets Using VizStruct, A Novel Multidimensional Visualization Technique

Kavitha Bhasi,¹ Li Zhang,² Aidong Zhang,² and Murali Ramanathan^{1,3}

Received December 3, 2003; accepted February 9, 2004

Purpose. Data visualization techniques for the pharmaceutical sciences have not been extensively investigated. The purpose of this study was to evaluate the usefulness of VizStruct, a multidimensional visualization tool, for applications in pharmacokinetics, pharmacodynamics, and pharmacogenomics.

Methods. The VizStruct tool uses the first harmonic of the discrete Fourier transform to map multidimensional data to two dimensions for visualization. The mapping was used to visualize several published pharmacokinetic, pharmacodynamic, and pharmacogenomic data sets. The VizStruct approach was evaluated using simulated population pharmacokinetics data sets, the data from Dalen and colleagues (*Clin. Pharmacol. Ther.* **63**:444–452, 1998) on the kinetics of nortriptyline and its 10-hydroxy-nortriptyline metabolite in subjects with differing number of copies of the CYP2D6, and the gene expression profiling data of Bohen and colleagues (*Proc. Natl. Acad. Sci. USA* **100**:1926–1930, 2003) on follicular lymphoma patients responsive and nonresponsive to rituximab.

Results. The VizStruct mapping preserves the key characteristics of multidimensional data in two dimensions in a manner that facilitates visualization. The mapping is computationally efficient and can be used for cluster detection and class prediction in pharmaceutical data sets. The VizStruct visualization succinctly summarized the salient similarities and differences in the nortriptyline and 10-hydroxy-nortriptyline pharmacokinetic profiles in subjects with increasing number of CYP2D6 gene copies. In the simulated population pharmacokinetic data sets, it was capable of discriminating the subtle differences between pharmacokinetic profiles derived from 1- and 2-compartment models with the same area under the curve. The two-dimensional VizStruct mapping computed from a subset of 102 informative genes from the Bohen and colleagues data set effectively separated the rituximab responder, rituximab nonresponder, and control subject groups.

Conclusions. The VizStruct approach is a computationally efficient and effective approach for visualizing complex, multidimensional data sets. It could have many useful applications in the pharmaceutical sciences.

KEY WORDS: microarray; pharmacodynamics; pharmacogenomic modeling; pharmacokinetics; visualization algorithms.

INTRODUCTION

Many pharmaceutical applications result in large multidimensional data sets. With the advent of genomic technolo-

gies such as DNA arrays and proteomics, which are now capable of simultaneously measuring the expression of thousands of genes from single samples, the data sets have become even larger.

Data visualization, however, has not been extensively investigated in the context of the pharmaceutical sciences. Visualization, using graphs and other techniques, is an efficient aid for finding structures, features, patterns, and relationships in a data set. However, multidimensional data sets, which arise in many pharmaceutical applications, present special visualization challenges because they cannot be easily represented in interpretable two-dimensional or three-dimensional graph formats. Good interactive, multidimensional visualization tools can provide additional perspectives that assist the user to understand large complex data sets at an intuitive level, facilitate subsequent hypothesis generation, and enhance the data mining experience. In this report, we investigate a visualization tool, VizStruct, which is capable of mapping multidimensional data to two dimensions, and assess its usefulness for several key pharmaceutical applications.

DERIVATIONS AND RESULTS

The Mapping

VizStruct is a projection that maps the n -dimensional vectors in the input data to two-dimensional points (3–5). If the vector $\mathbf{x}[n] = (x[0], x[1], \dots, x[n-1])$, represents a data item in n -dimensional space, \mathbf{R}^n , its mapping to a point $F_1(\mathbf{x}[n])$ in the complex plane C is given by:

$$F_1(\mathbf{x}[n]) = \sum_{j=0}^{n-1} x[j] e^{-2\pi i j/n} \quad (1)$$

In equation above, $i = \sqrt{-1}$ and the complex exponential has the effect of dividing the circle of display into equally spaced sectors. The real and imaginary components of $F_1(\mathbf{x}[n])$ are used for creating the two-dimensional mapping. Because the real and imaginary components representation of a complex number is mathematically equivalent to its amplitude (R) and phase (ϕ) representation, the VizStruct mapping can be plotted either on a Cartesian real and imaginary axis plot or on a polar plot; we use both plots in this report. The mapping $F_1(\mathbf{x}[n])$ is equivalent to the first harmonic of the discrete Fourier transform (DFT), which allows the fast Fourier transform algorithm to be used for computation. The VizStruct mapping is equivalent to the geometric projection technique used in radial visualization algorithms (3,4); however, our use of the first Fourier harmonic, which facilitates computation and provides insights into the underlying properties of the mapping, is novel (5).

The VizStruct mapping preserves the correlation relationships between vectors in the input and the output spaces. For example, all vectors of the form (a, a, \dots, a) are mapped to the center of the unit circle. If two vectors \mathbf{X} and \mathbf{Y} share geometric similarity, that is, $\mathbf{Y} = a\mathbf{X}$ where a is scalar, they will map to a radial line. All vectors with the same pattern will map to the same line, and those that are similar map to a narrow region around the line.

¹ Department of Pharmaceutical Sciences, State University of New York at Buffalo, Buffalo, New York 14260, USA.

² Department of Computer Science, State University of New York at Buffalo, Buffalo, New York 14260, USA.

³ To whom correspondence should be addressed. (e-mail murali@acsu.buffalo.edu)

Application of the VizStruct Visualization Approach in Pharmacogenetics

As a pharmacogenetics case study, we used the data from Dalen *et al.*, who demonstrated that the plasma concentration-time profiles of the antidepressant, nortriptyline, and its metabolite, 10-hydroxynortriptyline, were dependent on the number of copies of the cytochrome P450 2D6 isoform (CYP 2D6), which is involved in its metabolism (1). The subjects either had no functional copies or 1, 2, 3 or, in one case, 13 copies of the CYP 2D6 gene. The findings, summarized in Figs. 1A and 1B, show that individuals with more CYP 2D6 copies metabolize the drug more extensively, and the resultant metabolite levels are higher compared to subjects with few copies.

The polar plots in Figs. 1C and 1D are obtained using the VizStruct mapping. The results in Figs. 1C and 1D show that VizStruct succinctly captures the essential aspects of the kinetics in Figs. 1A and 1B, respectively. The inverse relationship between the kinetics of nortriptyline and its metabolite (subjects with more CYP 2D6 genes produce lower nortriptyline levels and higher levels of 10-hydroxynortriptyline) are represented by the mappings in Figs. 1C and 1D. The subjects with two and three copies of CYP 2D6 show similar nortriptyline kinetics in Fig. 1A and are mapped close to one another by VizStruct in Fig. 1C. The geometric similarity of the kinetic profiles is reflected in the radial arrangement of points in Figs. 1C and 1D.

These findings demonstrate the feasibility of using VizStruct as an intuitive visualization approach because it is capable of representing key features of multidimensional data sets succinctly in two dimensions.

Application of the VizStruct Visualization Approach in Population Pharmacokinetics

We reasoned that novel visualization tools could have a significant impact on population pharmacokinetics (PK) stud-

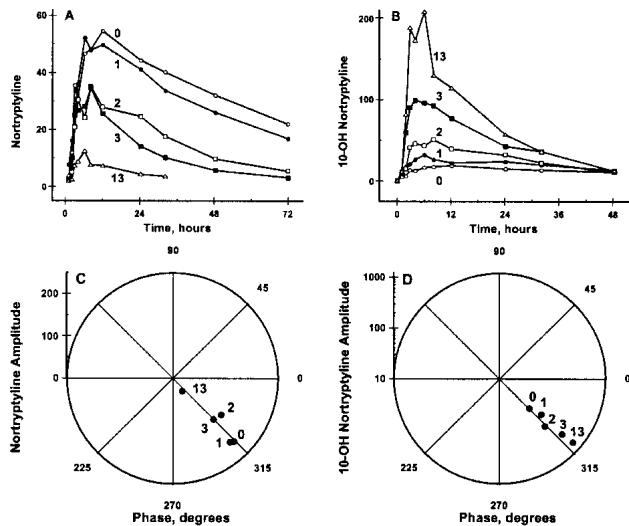


Fig. 1. (A) and (B) show the kinetics of nortriptyline and its metabolite, 10-hydroxynortriptyline, in subjects with 0, 1, 2, 3, or 13 copies of the CYP 2D6 gene (1). The drug and metabolite concentrations are in nM. (C) and (D) are the polar plot representations of the VizStruct mapping. The amplitude and phase of the first Fourier harmonic are shown. The numbers shown against the curves (A and B) and the points (in C and D) are the number of CYP 2D6 copies.

ies because parallel coordinates representations of drug concentration-time profiles get cluttered and difficult to interpret for large PK data sets. Here, we present results from a population PK simulation case study, in which two groups differing in clearance (clearance values of 5 volume units/time, and 7.5 volume units/time) were generated. For both groups, the dose and the volume of distribution of the central compartment were set to 100 mass units and 5 volume units, respectively. A clearance-parameterized, one-compartment model was used, and the ADAPT pharmacokinetic and pharmacodynamics systems analysis program (6) was used to simulate 100 individual PK profiles for each group assuming a log-normal distribution for the clearance and volume of distribution of the central compartment.

Figure 2A shows that the two groups differing in clearance are well separated by the mapping from their sampled kinetic profiles. The mean values of the “amplitude” are greater for the low clearance group because their mean concentrations are greater.

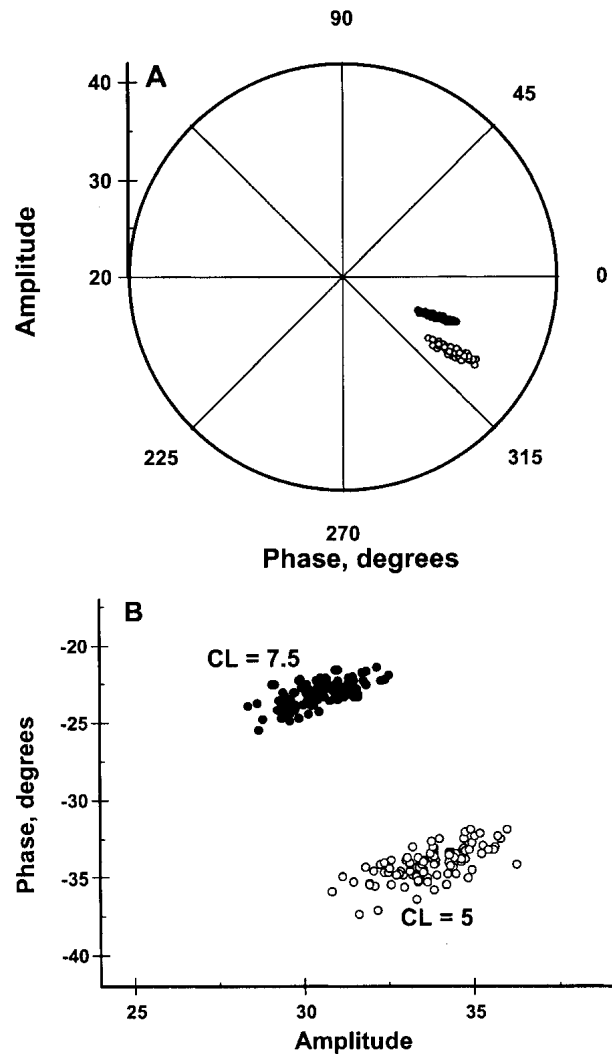


Fig. 2. Two different views, (A) a polar plot and (B) a linear plot of a simulated population pharmacokinetics data set. A one-compartment model was used: the CL was set to 7.5 volume units/time for one group and 5.0 volume units/time for the other; the volume of distribution was 5 volume units. A log-normal distribution was assumed for clearance and volume of distribution.

To further challenge the capabilities of VizStruct, we simulated the challenging PK data set shown in Figs. 3A and 3B. The curve shown in open circles is the simulation for a one-compartment model, and the remaining four curves are simulations from two-compartment models differing slightly in the A , α and B , β values of the bi-exponential relationship

$C = Ae^{-\alpha t} + Be^{-\beta t}$; all the curves have the same area under the curve ($AUC_{0 \rightarrow \infty}$), which causes the curves to overlap significantly. For the one-compartment model, the dose was set to 100 mass units; the volume of distribution was set to 5 volume units; the clearance was set to 10 volume/time units. Together, these parameters result in a degradation rate constant (K) of 2 time^{-1} , an initial concentration of 20 mass/volume units, and in a $AUC_{0 \rightarrow \infty}$ of 10 mass-time/volume units. The two-compartment simulations in Fig. 3 all had $AUC_{0 \rightarrow \infty}$ of 10 mass-time/volume units. The A and B , in the bi-exponential relationship were, respectively, set to 5 and 10 mass/volume units for all the curves. The α were 1, 2, 2.5, and 3 time^{-1} units with corresponding β values of 2, 1.33, 1.25, and 1.20 time^{-1} units, for the curves marked, respectively, with filled circles, open squares, filled squares, and open triangles that are discernible in Fig. 3B. Figure 3A shows that on linear axes, the data points are virtually superimposed on one another; the differences become apparent only in Fig. 3B because of the use of the logarithmic axes. Figure 3C shows that the four PK profiles are well separated in the VizStruct mapping. The two-compartment profiles are clustered with each other and well separated from the one-compartment profile. These findings suggest that VizStruct visualization may prove useful for exploring large pharmacokinetic data sets.

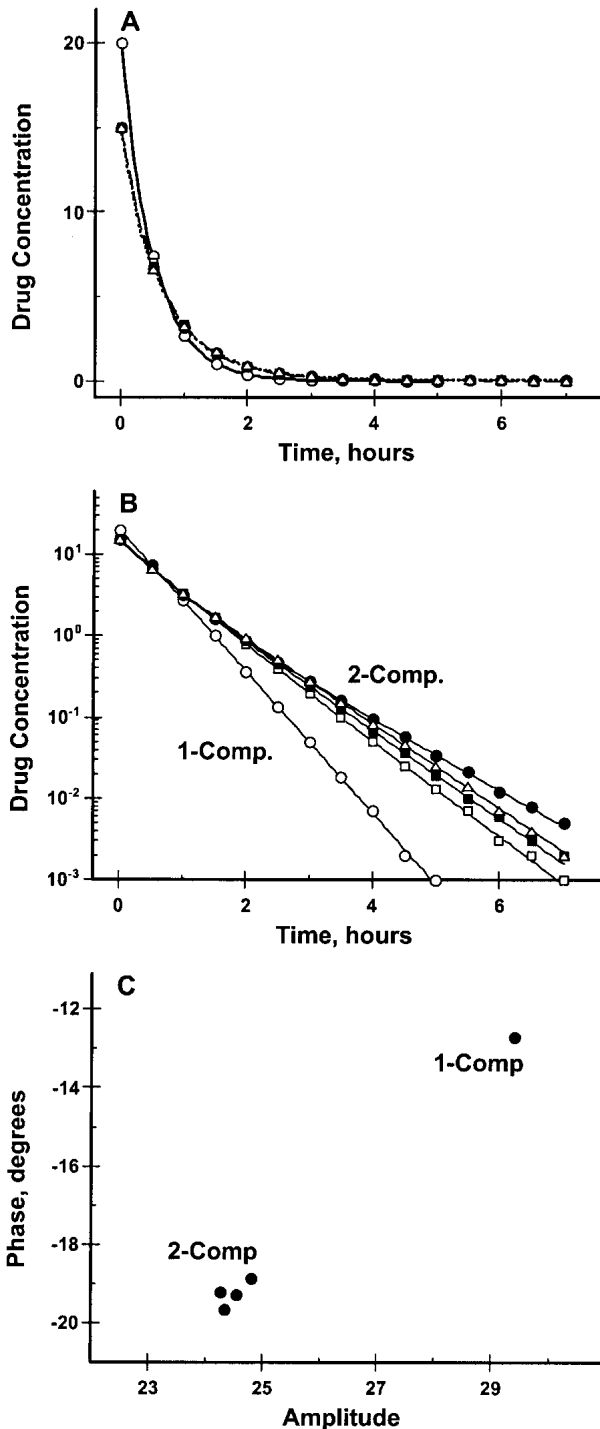


Fig. 3. (A) shows the simulated data from a one-compartment model (open circles) and three different two-compartment simulations (square, triangle, filled circle). (B) shows the same data as (A) on logarithmic y-axis. In (A) and (B), the concentrations are in arbitrary units. (C) shows the VizStruct mapping of the PK data set from (A).

Application of the VizStruct in Assessing Treatment Responses from Gene Expression Profiles

Gene expression profiling with microarrays provides simultaneous measurements of thousands of messenger RNAs and yields large multidimensional data sets that are a challenge to visualize. Bohen *et al.* measured the pretreatment gene expression profiles in follicular lymphomas of patients receiving rituximab, a monoclonal antibody directed against the CD20 protein found on B cells (2). These authors suggested that the gene expression patterns of lymphomas that were nonresponsive to rituximab therapy were more similar to those of control lymphoid tissues than lymphomas that responded to therapy. We used the significance analysis of microarrays (SAM) algorithm (7) to identify an informative subset of 102 genes from the data set and projected the 102-dimensional vectors for each sample to 2 dimensions using the VizStruct algorithm (Fig. 4). The VizStruct projection separated the partial/complete responders (open circles) from the nonresponders (filled circles) and normal controls (triangles). One nonresponder sample was not separated from the partial/complete responder cluster. Additionally, the proximity and radial location of the nonresponder cluster relative to normal controls indicate that the nonresponder group was somewhat more similar to the normal control group than the partial/complete responder group. Thus, the findings obtained through visualization are generally consistent with the conclusions of Bohen *et al.* (2).

DISCUSSION

In this report, we presented case studies for several pharmaceutical applications ranging from pharmacogenetics, population pharmacokinetics, and pharmacogenomics for which the VizStruct multidimensional visualization approach can prove useful.

The pharmacogenetics case study of the Dalen *et al.* data

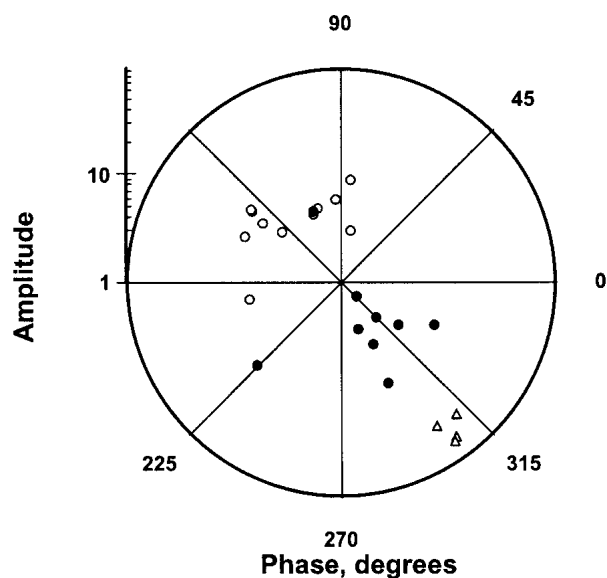


Fig. 4. This figure shows the results from projecting the expression levels of the 102-gene subset of informative genes from the data of Bohen *et al.* (2). The 11 partial/complete responder samples are shown in open circles; the nonresponders are shown in filled circles, and the 4 normal tissues are shown with the open triangles.

set was kept intentionally simple so that the findings from VizStruct could be easily compared to the interpretations obtained by visual inspection of the time profiles. Therefore, this example should be considered as a feasibility study rather than a rigorous demonstration of VizStruct capabilities because it could be argued that the original data from Dalen *et al.* could be visualized simply by plotting the t_{\max} (time at which the maximal concentration C_{\max} occurs) against the C_{\max} , and that the use of VizStruct was unnecessary. To further challenge the capabilities of VizStruct, we examined other, more difficult data sets in the remaining case studies.

We examined two pharmacokinetic data sets that were structured using simulations but included more difficulty than the pharmacogenetics case study. In pharmacokinetics, exponential decay processes are commonly used for modeling, and VizStruct has an excellent ability to succinctly represent PK profiles with a wide range of half-lives. For example, the unit bolus—which can be viewed as an extremely fast, first-order process—is mapped to amplitude $R = 1$, and phase $\phi = 0$, while a constant function—which does not change with time and can therefore be viewed as an extremely slow, first-order process—is mapped to $R = 0$ and $\phi = 0$. VizStruct is also relatively insensitive to noise because the discrete Fourier-transform sum-product, which underlies the mapping, reduces the effects of noise.

The case study of the gene expression profiles reported by Bohen *et al.*, which had 102 genes, highlights the ability of VizStruct to handle large multidimensional data sets (2). Despite the large number of genes and simplicity of the visualization, we found that the VizStruct approach was capable of recapitulating the key findings of Bohen *et al.*, which were arrived at using a hierarchical clustering algorithm.

The VizStruct approach differs distinctively from competing visualization methods such as parallel coordinates and multidimensional scaling (MDS). Most pharmaceutical scientists use the parallel coordinates approach for visualizing the

data; that is, each dimension is plotted along a separate axis. The parallel coordinates approach gets easily cluttered and increasingly difficult to interpret when used with large data sets with modest levels of noise. The MDS approach is the current gold standard for multidimensional visualization. In MDS, the presentation in two dimensions is optimized to preserve a specific aspect of the relationship, for example, the Euclidean distance, block distance, or rank relationships between the points in the N -dimensional space. In practice, MDS encompasses a class of methods, depending on the stress function being optimized. For many pharmaceutical applications, Sammon's nonlinear mapping is probably the most appropriate MDS method because it normalizes the distances in the stress function to distances in the original N -dimensional space. Generally, Sammon's mapping improves visualization of data sets containing a wide dynamic range of values. Despite providing results that are mathematically optimal in some sense, MDS (Sammon's mapping included) is not ideal primarily because the incremental addition of even a single point requires a complete repetition of the optimization procedure and possible extensive reorganization of all the previously mapped points to new locations. It is computationally intensive with time complexity of $O(n^2)$, where n is the number of points, because of the time-consuming function evaluations and iterations required for optimization. VizStruct, in contrast, is computationally efficient with time complexity of $O(n \log n)$ and does not require any rearrangements to accommodate incremental points.

In conclusion, VizStruct is effective, flexible and versatile. The method is also computationally efficient and has sound theoretical underpinnings. It may prove a useful approach for multidimensional data visualization in many pharmaceutical sciences applications.

ACKNOWLEDGMENTS

This work was supported in part by Grant RG3258A2 from the National Multiple Sclerosis Society. Support from the National Science Foundation (Research Grant 0234895) and the National Institutes of Health (P20-GM 067650) is also gratefully acknowledged.

REFERENCES

1. P. Dalen, M. L. Dahl, M. L. Ruiz, J. Nordin, and L. Bertilsson. 10-Hydroxylation of nortriptyline in white persons with 0, 1, 2, 3, and 13 functional CYP2D6 genes. *Clin. Pharmacol. Ther.* **63**:444–452 (1998).
2. S. P. Bohen, O. G. Troyanskaya, O. Alter, R. Warnke, D. Botstein, P. O. Brown, and R. Levy. Variation in gene expression patterns in follicular lymphoma and the response to rituximab. *Proc. Natl. Acad. Sci. USA* **100**:1926–1930 (2003).
3. D. Bhadra. Masters thesis. An interactive visual framework for detecting clusters of a multidimensional dataset. State University of New York at Buffalo, 2001.
4. P. Hoffman, G. Grinstein, K. Marx, I. Grosse, and E. Stanley. DNA visual and analytic data mining. In *Proceedings of the 8th IEEE Visualization '97 Conference*, Phoenix, AZ, 1997, pp. 437–441. IEEE Computer Society, Washington, DC.
5. L. Zhang, A. Zhang, and M. Ramanathan. VizStruct: exploratory visualization for gene expression profiling. *Bioinformatics* **20**:85–92 (2004).
6. D. Z. D'Argenio and A. Schlumitzky. *Users Guide to Release 4: Adapt II Pharmacokinetic/pharmacodynamic systems analysis software*. Biomedical Simulations Resource. University of Southern California, Los Angeles, CA, 1997.
7. V. G. Tusher, R. Tibshirani, and G. Chu. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA* **98**:5116–5121 (2001).