Clustering of CD Spectral Data as a Prototype QSAR Model for Neuropeptides

NEIL PURDIE* AND DENNIS W. PROVINCE

Contribution from *Department of Chemistry, Oklahoma State University, Stillwater, Oklahoma 74078-0447*

Received June 25, 1999. Accepted for publication September 23, 1999.

Abstract □ An analytical method that might eventually qualify as a general quality control assay procedure for polypeptide drug forms was described in the companion article to this paper. The detector is visible range circular dichroism spectroscopy. Multivariate data analysis reduced the spectral data to essentially four principal components (or factors) that are characteristic of each analyte. The level of analytical selectivity achieved among 51 analytes is very high. Using an alternative factor analysis algorithm, the selectivity is even more conveniently accomplished in the form of a 2-D cluster diagram presentation that has the potential of being a prototypical predictive in vitro model for correlating experimental data with structure–activity or structure–function relationships. Clustering of the analytes is a consequence not only of the chiral interactions associated with ligand exchange in the immediate primary coordination sphere of the host derivatizing reagent, but also of long-range intermolecular interactions between the coordination architecture of the host and the chiral polypeptides.

Introduction

Chirality is a common molecular property among natural products and an important factor in determining the efficacy of drug substances.[1] For an enantiomeric pair the isomer with proven clinical therapeutic value is called the eutomer, the other the distomer. Possibilities exist that a distomer can be toxic or mutagenic, making the assay of both enantiomers an essential part of regulatory tests. Measuring the chemical purity of an enantiomeric form is done by conventional standardized methods. Accurate measurement of the enantiomeric purity (EP) is a much more difficult problem. How to measure EP's quantitatively and routinely has recently become a major concern for manufacturing and regulatory agencies. The measurement of EP, or the related quantity enantiomeric excess (EE), expressed as a ratio of the two isomers, has now reached the status of being a requirement imposed by the Food and Drug Administration in new drug submissions.

Manufacturers have the option of choosing between either the enantiomeric resolution of readily synthesized racemic mixtures, or stereospecific syntheses.[1] Resolutions are typically long, demanding, and costly processes which in the end invariably fail to achieve 100% EP. Chiral chromatographic procedures significantly reduce the work load and costs for enantiomeric resolutions,[2,3] but separations that are complete are still not generally accomplished. More and more often, stereospecific synthesis has become the favored option. Since 100% EP is seldom achieved, both procedures are perhaps better thought of as enantiomeric enrichment processes. Monitoring the possible reduction of EP during the shelf life of a drug substance through racemization, especially if it is in solution or in suspension, is an analytical add-on to protocols that are submitted for approvals.

As a result, regulatory agencies are facing a rapidly increasing work load with respect to certifying and approving analytical methods to measure not only chemical purities but also EE (impurities) of chiral drug forms.[4,5] There is also the issue of defining what are the acceptably safe EE threshold levels below which a product is to be rejected from either manufacture or distribution. To ease the load, there is an urgent need for accurate and convenient routine experimental procedures that will give chemical and chiral impurity data, and preferably in an automated way if at all possible.

The problem is immense and still expanding. Pharmaceutical and biotechnology industries are investing more and more R&D resources into the production of chiral substances. With the advent and continued momentum of combinatorial chemistry,[6–9] the number of potential new chiral drug substances is increasing exponentially. Nowhere, perhaps, are these numbers increasing faster than in the area of peptide, peptidomimetic, and protein drug forms. Identifying which real or virtual compounds are worthy of further development is an exceptionally difficult problem. Culling out the few new potential drug substances by performing individual in vivo assays was not really an option and led quickly to the development of mixture assays.[7–9]

Viable alternatives, with the potential for automation, are quantitative structure−activity relationship (QSAR) screening algorithms that are based upon in vitro chromatographic or spectroscopic data of one kind or another. Exploitation of QSAR models is done in two parts, calibration and prediction. In the first step, measured experimental data and therapeutic properties of known drug forms are mathematically correlated to derive a calibration model. The calibration model is in turn exploited to predict the therapeutic potential of a new drug form. The prediction is based entirely on the measured experimental data, e.g., retention times or absorbance, for each new substance. The broader the therapeutic range represented in the calibration, the better the prediction model. These particular experimental methods, however, do not always address the question of how important molecular chirality is, in the therapeutic function of a drug.

The motivation for this study and previous related studies on peptides and proteins[10–13] was to address the lack of emphasis that is given to causal relationships between chirality and function, especially since molecular chirality is so commonplace among natural and synthetic drug substances. Protein receptor sites are chiral. How the chirality of a receptor kinetically reconfigures itself to accommodate the chirality of the drug molecule, and vice versa, is a factor in the discrimination between eutomers and distomers. That steric selectivity has been exploited, almost without recognition, in the development of a myriad of clinical assays that are based upon analytical specificities attributed to monoclonal and polyclonal antibodies and their interactions with antigens. Despite the importance

10.1021/js990210d CCC: $18.00
Published on Web 10/28/1999

of the chirality recognition, at no time has a test been based upon chirality detection. Antibodies have been elevated to the exalted rank of Analyte Specific Reagents (ASR). That same high selectivity, however, is a natural deterrent to the derivation of broad based or generally predictive QSAR models.

Direct spectroscopic methods are a more attractive option over chromatography because of their comparative simplicity of operation and consistent reproducibility in the recalibration of standards, etc. Mass spectrometry is selective but not with respect to chirality differentiation. Detectors that measure absorbance and fluorescence in the visible-UV range are notorious for their lack of analytical selectivity, but can be improved upon by using derivatization reactions, multivariate analyses methods, such as principal component analysis (PCA), factor analysis, and artificial neural networking in data handling,[14] or experimentally by introducing time as an added analytical variable.

Spectroscopic methods with the power to detect molecular chiralities are circular dichroism (CD) and optical rotatory dispersion (ORD). In the specialized area of screening peptides and proteins, however, the vast majority of CD investigative work is done on data from the UV range.[15] The purposes there are to obtain predictive information that can be used to resolve the question of 3-D structural properties of proteins in solution, apportion solution structures between helix and sheet forms, and investigate the kinetics and thermodynamics of folding–unfolding mechanisms, etc. CD bands in the UV are broad and without discrete features. Like absorbance, their potential for establishing QSAR models is limited.

In a series of recent articles, visible range CD spectral data were used to selectively differentiate among di- and tripeptides,[10,11] insulins,[12] and neuropeptides.[13] Mathematical algorithms were derived with which to determine their chemical and enantiomeric purities.[12,13] The absorbance property required for CD to occur in the visible range, was provided by first binding the peptides to Cu(II) ion in pH 13 aqueous solution. Visible absorbance spectra of Cu(II) complexes are comprised of three electronic transitions[16] all of which are CD-active. For the completed studies, peptides were either bound directly to the copper ion[10,11] or displaced an auxiliary ligand,[12,13] e.g., D-histidine. Major accomplishments from these articles that are of relevance to the goals of this one are the ability to discriminate between human and porcine insulins[12] (two 51 amino acid residue proteins whose sequences differ in the identity of one acid at the B30 terminal amino acid only) and between human insulin and the human LysPro variant form in which the order of the B29 L-lysine and B30 L-proline are reversed, and to achieve a level of *quantitative* analytical selectivity that approaches specificity among 51 peptides.[13]

The specific objective of this article was to derive a cluster correlation diagram between visible range CD data on one hand and molecular stuctures, and their corresponding clinical therapeutic functions, on the other. Success would mean we will have a prototypical in vivo predictive QSAR model of some general applicability. CD spectral data are taken from the previous companion article.[13] Only tripeptides and longer oligopeptides are included to test the general practicality of a clustering algorithm. Special mention is given to the subgroup of 19 neuropeptides, classified by their particular clinical functional actions.

Experimental Section

Chemicals—Thirty-five oligo-and polypeptides used for the study were the tripeptides GGA, GGG, GGH, GGI, GGL, GGF, GHG, LGG, and YGG; the higher oligoglycines GGGG, GGGGG, and GGG-GGG; the enkephalins DSLET, DTLET, DADLE, DAGO, DALDA amide, DPDPE amide, Met^5-enkephalin, Leu^5-enkephalin, Leu^5-enkephalin amide, (D)-Ala^2-Leu^5-enkephalin amide, and β-endorphin; the Dynorphins A (1–9), A (1–11), A (1–13), A (1–13) amide and B (1–13); and the miscellaneous CTAP amide, ICI 174,-864, PLO 17 amide, and human, porcine, Lyspro, and bovine insulins. Amino acid sequences are included in the footnote to Table 1.

Commercial sources for the oligopeptides and insulins were listed previously.[13] Reagent grade D-histidine, reported to have an EP better than 99.8%, was obtained from Sigma Chemical Co. Reagent grade $CuSO_4 \cdot 5H_2O$ was obtained from Fisher Scientific.

Solution Preparations—These are described in greater detail in the prior companion article on the 51 peptide and protein forms.[12,13] Working solutions for the assays were prepared by diluting aliquots of Cu(II)–D-histidine stock solutions by a factor of 10 with 0.10 M NaOH such that the final composition was 2.0 mM in Cu(II) ion, 8.0 mM in D-histidine, and 3.0 mM in KI added as stabilizer, in 0.1 M base. To make future quality control (QC) procedures more amenable to automation, we chose to add equal masses of the analytes (10 mg) to the aliquot of stock prior to its dilution rather than use equal concentrations. From the range of molar masses of the neuropeptides, the actual analyte concentrations in the working solutions varied from 0.2 to 1.20 mM. All are intentionally lower than the total copper ion concentration and not enough to exchange completely with the D-histidine ligand. With modern CD instrumentation, the mass of analyte can be reduced 10–100 fold.

Measurements—CD spectra were measured using a Jasco 500-A automatic recording spectropolarimeter coupled to an IBM-compatible PC through a Jasco IF-500 II serial interface and data processing software. Experimental parameters were wavelength range 400–700 nm, sensitivity 100 mdeg/cm, time constant 0.25 s, scan rate 200 nm/min, path length 5.0 cm, temperature ambient. With 2 nm spectral resolution, the full measured spectrum consists of 1500 data points.

Calibrating the day to day reproduciblity of the system was done by measuring the CD spectrum for the Cu(II)–D-histidine reference working solution. Statistical data used to determine spectral reproducibilities were based on the standard deviations (SD) for the maximum ellipticities measured at the wavelengths 487 and 682 nm. The SD values were 7.42 ± 0.07 mdeg and -214 ± 0.60 mdeg, respectively, within and between stocks.[10,11]

Results and Discussion

The assay works on the premise that chiral ligand exchange of the polypeptides for D-histidine, complexed to Cu(II) ion in strong aqueous base solution, will induce a change from the CD spectrum for the host that is characteristic of the peptide analyte. The host metal complex serves as a CD-inducing and color-derivatizing agent for the neuropeptides which otherwise would be CD transparent in the visible range. Spectra for the products,[13] uncorrected for the host spectrum, are the bases for subsequent mathematical correlations.

If there is any truth to the concept that the CD spectra for the mixed complexes correlate with molecular, and therefore therapeutic, function, there should be strong spectral similarities within specific groups and obvious dissimilarities among groups. In other words the focus is on proving a pattern recognition of polypeptides by type. With this thought in mind, we can dispense with the conventional practice of measuring metal–ligand stoichiometries which should be similar within each subset, and dispense with the need to measure formation constants, expecting the variations in these values within a group to vary very little. Therefore, provided the specified procedure for the assay is defined and followed in every detail, analyte recognition and the measured accuracy in their determinations will be considerably improved over univariate (one wavelength) determinations.

The mathematical basis for achieving the analytical selectivity claimed in the previous article[13] was data

Table 1—Principal Component (PC) Values and Cluster Assignments Derived from the PCA Calculation Made from the Correlation of *Y*-Matrix for Tripeptides, Oligopeptides, and Insulins[a]

peptide	PC1	PC2	PC3	PC4	PC5	PC6	cluster
GGH	1.90754	1.07084	–5.80304	0.56910	0.76218	0.21884	1
GGI	–3.33190	0.93815	–1.31367	0.58205	0.24248	–0.20411	2
GGL	–2.48178	1.54899	–0.83902	0.41809	0.19503	0.00979	2
GGF	–7.02666	2.48361	–3.42084	1.26270	0.26464	–0.10426	2
GGA	–3.30146	3.77231	–1.58899	0.32102	0.18140	–0.05297	2
LGG	–1.08595	4.62576	0.36281	–0.11674	–0.31355	–0.04903	3
YGG	2.21496	5.92887	2.47355	0.39291	0.07784	–0.13505	3
GHG	0.71513	5.73447	2.83448	–1.50678	0.67706	0.58728	3
GGG	–1.00748	4.63183	–0.13142	–0.38407	–0.38924	0.05335	3
GGGG	–0.98249	5.10671	–0.20555	–0.41400	–0.41839	0.05011	3
GGGGG	–0.99244	5.08708	–0.19114	–0.44882	–0.43975	0.02403	3
G6	–0.77889	3.13769	0.20912	–0.37229	–0.29460	0.05897	3
DSLET	5.61386	–1.78747	0.05173	–0.40400	–0.03030	–0.07857	5
DTLET	7.31482	–1.43674	0.18100	–0.71373	0.16854	–0.28908	5
DADLE	6.56303	–1.40095	–0.08242	–0.55754	0.03430	0.08077	5
DAGO	3.42481	2.26752	3.32135	1.53333	0.39364	–0.18805	3
DynA9	–2.14370	–2.52980	1.56920	0.06199	–0.31105	–0.06093	4
DynA11	–1.76459	–1.76269	2.87524	0.67279	–0.10279	0.16660	4
DynA13	–1.68912	–2.25739	2.67589	0.60244	–0.03417	0.16794	4
DynAA13	–3.56666	–1.56817	2.52575	0.30886	–0.11321	0.02043	4
Met-enk	–0.23515	–0.57906	–0.51375	0.22497	0.00360	–0.01722	1
DynB13	–0.85486	–2.64485	1.81267	–0.06927	–0.33628	0.04186	4
β-endorph	–1.99505	–0.50226	2.52819	0.28876	0.38610	–0.01400	4
DALDA	2.65449	–4.29649	–2.52405	2.10245	–0.56252	0.57465	1
CTAP	–8.64881	–4.24805	–3.13553	–2.15925	–0.19259	0.10251	6
DPDPE	8.93870	–0.60145	–2.26189	–0.70663	0.34742	–0.03321	5
ICI174	0.05242	–5.16070	1.98254	–0.22740	0.19753	–0.04955	4
PLO17	0.09633	–5.23027	1.97495	–0.21865	0.22239	–0.00471	4
α-Leu-enk	7.96527	–1.32821	–1.88627	0.03077	–0.34296	0.02164	5
L-enkeph	2.32657	0.48518	–1.23225	0.05665	–0.48166	–0.18311	1
Lenkamid	2.83518	0.95343	–0.72967	–0.17806	–0.45585	–0.21976	1
human ins	–2.64882	–2.89625	–0.41242	–0.34787	0.11478	–0.12309	4
porcine	–2.72926	–2.93333	–0.46651	–0.35625	0.14004	–0.10869	4
lyspro ins	–2.87639	–2.22205	–0.51801	–0.00418	0.23594	–0.04509	4
bovine ins	–2.48166	–2.38724	–0.02859	–0.23897	0.17400	–0.1182	4

[a] Key to Structures reading from amine end. Lower case letters are D-enantiomer forms, (asterisks indicate ring structures): DSLET (YsGFLT); DTLET (YtGFLT); DADLE (YaGFL); DAGO (YaG(N-Me)FG); DALDA (YrFK amide); DPDPE (Ype*GFpe*); CTAP (fC*Y-wRTP*eT amide); DynorphinA (1–9) (YGGFLRRIR); DynorphinA (1–11) (YGGFLRRIRPK); DynorphinA (1–13) (YGGFLRRIRPKLK); Dynorphin B (1–13) (YGGFLRRQFKVVT); b-endorphin (YGGFMTSEKSQTPLVTLFKNAIKNAYKKGE); Met5-enkephalin (YGGF(N-Me)M); Leu5-enkephalin (YGGFL); (D)-alanine-Leu5-enkephalin (YaGFL); ICI 174,864 (N,N-diallyl YAiAiFL);

reduction of the 1500 data points in the CD spectra for each analyte, taken separately, to twelve factors (3 eigenvalues and 9 eigenvectors) using PCA done on Spinning Plot data.[17] Spinning Plots are created by plotting wavelength, CD spectral data for the Cu(II)–D-histidine host complex, and the Cu(II)–(D-histidine-peptide) mixed complexes on the *x*, *y*, and *z* coordinate axes, respectively. Calculations are described in detail in the prior article.[13] Factors showing the most sensitivity to the identities of the analytes are the eigenvectors. The real analytical value of the tabulated data is the quantitative nature of the chemical and EP assays that can be accomplished using PCA data vs concentration correlations. Potential correlations of principal components with molecular structures are difficult to find when data are in tabular form.

PCA and Clustering Model—The model was developed using only the spectral data for the tripeptides and higher oligomers. Dipeptides were excluded because they lack the 3-D rotational conformational structures that are common to longer peptides.

In clustering, the strategy is to base correlations upon how well spectral patterns align with molecular "types" in a graphical manner. The major departure from Spinning Plot data reduction algorithm is that the spectral data for all 35 analytes are treated collectively. What follows is a description of the genesis of a possible calibration model. Spectral data for new analytes are required before the relative success of the model in making predictions can be tested.

To make modeling calculations more managable for a desktop computer, the original 1500 data points were reduced in number to 31 by choosing values at 10 nm intervals. The first step was to group the (35 × 31) CD data sets through a (correlation-of-*Y* variables) calculation, that expresses the relative strengths of successive pairwise correlations of CD signals in the form of a square matrix. PCA was then used to reduce the dimensionality of the correlation-of-*Y*'s matrix to just six factors or principal components (PC's), Table 1. These six factors account for 99.93% of the total variations among all of the *Y* variables in all 35 individual spectra. The first factor, PC1, has the maximum variation. The second factor, PC2, is orthogonal to the first and has the next greatest variation, and so on through all six factors.

Clustering is a technique of grouping together rows of PC's that share similar values across a number of variables. The hierarchical clustering option was used here. In this option, clustering starts with each point being its own "cluster". Through a series of iterative steps the two closest clusters are successively combined, by Ward's method, until ultimately all points are grouped as a single cluster. Between the extremes of 1 and 35 clusters there is an optimum number on which to base the model. As a rule of thumb the number of clusters would not exceed the number of "types" of analytes, assuming that that prior knowledge is available. Selection of the optimum cluster condition is linked to the relative magnitudes of the proportional changes between successive PC values. For this data set,

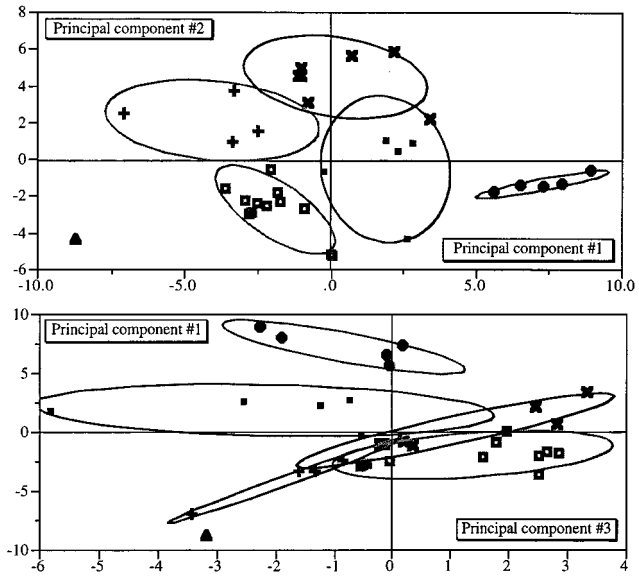

Figure 1—(a) Correlation of PC1 vs PC2 (Table 1) derived from the PCA of the correlation of *Y* matrix for the 35 peptides used for the model. Individual clusters are associated with (+) GGX tripeptides; (x) oligoglycinates and non-GGX tripeptides; (●) enkephalins; (◘) Dynorphins; (■) insulins (black); and (+) D^2-modified enkephalins. (b) Correlation of PC3 vs PC2 from the same source. The perspective is from a direction that is at right angles to the projection in part a. The same markers are used to designate the peptides by type. The special reason for part b is to demonstrate that the Dynorphins and the insulins are separable in the third dimension.

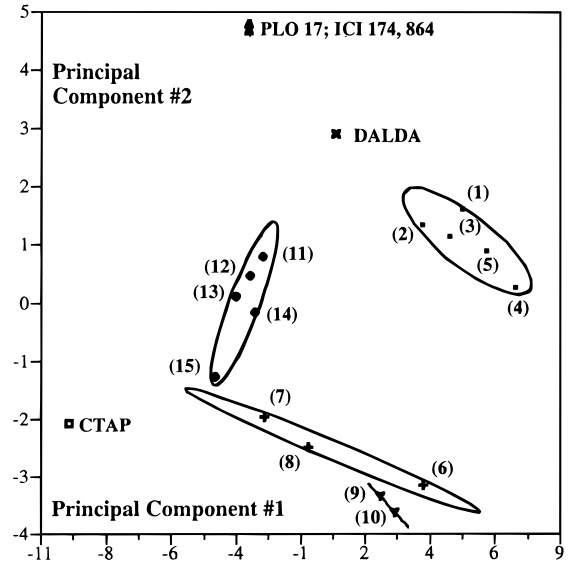

Figure 2—Repeat of the correlation of Figure 1a drawn exclusively for the 19 neuropeptides only. Cluster outliers are identified by name. Coordinates for the host D-histidine complex would correspond with the coordinates for PLO 17 and ICI 174,864. The δ-receptor cluster consists of (**1**) DTLET; (**2**) DSLET; (**3**) DADLE; (**4**) a^2-Leu^5-enkephalin; and (**5**) DPDPE. The μ-receptor cluster consists of (**6**) DAGO; (**7**) Met^5-enkephalin; and β-endorphin. The alternate δ-receptor cluster is comprised of (**9**) Leu^5-enkephalin and (**10**) Leu^5-enkephalin amide. The κ-cluster of the Dynorphins consists of (**11**) B (1–13); (**12**) A (1–13); (**13**) A (1–9); (**14**) A (1–11); and (**15**) A (1–13) amide.

the optimum number of clusters, or types, is 6; the tripeptides, oligoglycinates, dynorphins, enkephalins, insulins, and a miscellaneous group composed of the structurally unrelated CTAP, ICI 174,864, and PLO 17.

Figure 1 is a presentation of two 2-D cluster diagrams prepared by plotting different pairs of PC values, PC1 vs PC2 in Figure 1a, and PC3 vs PC1 in Figure 1b. The relationship between the two figures is that the projection of the lower diagram is from a direction that is orthogonal to the projection for the upper diagram. Together they represent two 2-D perspectives of a 3-D figure. Figure 1a shows the excellent segregation of the data points over five distinct cluster areas that encompass multiple points, and one "area" that is but a single point, i.e., CTAP. The four insulins and the Dynorphins appear to occupy a single cluster in the 2-D space of Figure 1a, but on viewing the same diagram from an orthogonal direction, Figure 1b, the types are clearly separated in the third dimension. Future studies should be able to exploit 3-D clustering option in a more direct manner for even greater discriminations.

The range in values for all PC's is very wide which enables good spatial separation of the clusters. Limits on the range of PC1 values are set by the Cys-Pen and Pen-Pen disulfide linked ring peptides, CTAP and DPDPE, which produce the greatest changes in the CD spectral data from that of the host complex.[13] Two other structural features of CTAP make it unique. It is the only peptide in the pool with an aromatic substituent on the amine terminus, and that residue is also in a D-enantiomeric form.

Neuropeptides Subgroup—Clustering of the 19 neuropeptides subgroup is shown separately from the tripeptides and insulins data in Figure 2. All except CTAP and ICI 174,864 have a Y-residue at the amine terminus. Data points on the PC1 vs PC2 plot are segregated over three distinct areas, plus a few individually distinct locations. The next question is whether these areas correlate with different protein receptor selectivities and affinities.

Three known peptide neuroreceptors, μ, δ, and κ, have been identified in the literature.[18,19] It has been proposed, but not generally agreed upon, that there is duality in the structures of potential μ and δ receptors. Neuropeptide drug forms have special affinity for one or another of the receptors. Dynorphins, for example, are structurally disposed to bind to κ receptors. Some drugs are reported to show multiplicity by binding to more than one receptor.

Beginning with the cluster outliers, CTAP is a confirmed μ-selective antagonist. DALDA on the other hand is considered to be the most selective μ-agonist known. Their relative locations are consistent with the diversity in their affinities. Distinction between PLO 17 and ICI 174,864 (a μ-agonist and a δ-selective antagonist, respectively), however, cannot be made since the spectral evidence is that they do not participate in ligand exchange. The host Cu(II)—D-histidine complex, therefore, would occupy the same coordinate position and be a natural reference point against which to "measure" relative receptor selectivities and affinities.

Compounds **1**—**5** in Figure 1 are the designer D^2-enantiomer-enkephalin analogues, DTLET, DSLET, DADLE, a^2-Leu^5-enkephalin, and DPDPE, respectively. These are recognized δ-receptor drugs. Structurally DADLE is the enantiomer of a^2-Leu^5-enkephalin because of the inversion of the chirality at the acid terminus. The chirality change appears to have little effect on their proximity in the cluster which might have been expected since it is located so far from the amine terminus which is the primary point contact between the ligand and the host metal complex. DALDA fits the description of a D^2-enantiomerically substituted enkephalin, yet it lies outside the "δ-receptor cluster". The result is consistent with the fact that DALDA favors binding to a μ-neuroreceptor (vide supra). Its absence from the "δ-receptor cluster" might be related to its being the only tetramer in the class.

The two natural Leu^5-enkephalins **9** and **10**, in which all residues are L-enantiomer forms, also show a preference for δ-receptors.[18,19] Their coordinates, however, are quite remote from the designer D^2-enkephalins "δ-receptor cluster". Effects of enantiomeric substitutions at amino acid

terminus residues may be a factor in proposing that there is structure multiplicity among neuroreceptors and among agonists.

DAGO (**6**) is yet another designer D^2-substituted enkephalin whose location is remote from the "δ-receptor cluster". Its spectral properties and PC coordinates place it in an entirely different cluster with Met^5-enkephalin (**8**) and β-endorphin (**7**) which are μ-receptors. DAGO is also classified as a μ-receptor agonist.[18,19] The locations of the Met^5 analogue, β-endorphin, and DAGO together define the "μ-receptor cluster" space. Structurally what DAGO and Met^5-enkephalin have in common is an N-Me substituent on the third and fourth residues, respectively. The substituent lies outside the plane of the primary tetragonal coordination sphere of the Cu(II) ion complex and as such adds a new chirality dimension to the ligand−ligand interactions. The observation that DAGO is a confirmed μ-receptor agonist and associates itself in space with two other μ-receptors, while its structure would suggest otherwise, is perhaps the most compelling reason for believing that a calibration model has been identified that could be the genesis of a QSAR model for predicting the receptor selectivities of new peptide drug forms.

As a general rule, Dynorphins bind to the κ-receptor.[18,19] All five included here form a tight single "κ-receptor cluster" which we suspect is related to their having identical initial sequences and a diminished outer-sphere involvement with the first coordination sphere of the metal complex. In close juxtaposition to the Dynorphins is the 30-residue β-endorphin which has the same sequence as the Dynorphins over the first four residues. Areas where the clusters impinge or intersect are conceivably indicators of possible duality in receptor binding functions.

The preliminary calibration model has provided some encouraging results. As more neuropeptides are added to the data pool, the boundaries and the compositions of the clusters will necessarily change. A much larger data pool is required before the predictive capabilities of the model can be critically and confidently tested.

Summary

An idea for a potential QSAR model for peptide drug substances is described. The model is based upon visible range CD data measured for a series of mixed Cu(II)−(D-histidine-peptide) complexes. Spectral responses to the ligand substitution reactions are a function of the identities of the coordinating bases in the analyte ligands and of the extended 3-D solution structure of the peptides. The model differs from other QSAR models in that longer range intermolecular forces that contribute to receptor−agonist interactions are a tangible part of the model. The value of the model as a predictive tool is still uncertain because the number of objects is still very small. It is quickly and easily expandable as other neuropeptides become available. The focus here was on neuropeptides. Drugs whose pharmacological functions are different will, in all likelihood, be treated as separate entities in cluster diagrams that are specific to each category or type.

References and Notes

1. Collins, A. N.; Sheldrake, G. N.; Crosby, J., Eds. *Chirality in Industry, The Commercial Manufacture and Applications of Optically Active Compounds*; Wiley: New York, 1994.
2. Fell, A. F., Ed. Special Issue on Chiral Discrimination. *Trends Anal. Chem*. **1993**, *12*, 125−189.
3. Ward, T. J.; Armstrong, D. W. In *Chromatographic Chiral Separations*, Zief, M., Crane, L. J., Eds.; Chromatographic Science Series, Vol. 40; Marcel Dekker: New York, 1988.
4. Rapaka, R. S.; Dhawan, B. N.; Renugopalakrishnan, V. An update of selected topics in the biology and chemistry of opiod peptides. *Opiod Peptides: An Update; NIDA Research Monograph 87*; National Institute on Drug Abuse: Washington, DC, 1988; pp 217−232.
5. Gueriguian, J. L.; Chiu, Y. Y. H. A few thoughts on the development and regulation of neuropeptides *Opiod Peptides: An Update; NIDA Research Monograph 87*; National Institute on Drug Abuse: Washington, DC, 1988; pp 405−413.
6. Houghten, R. A.; Dooley, C. T. Synthetic peptide combinatorial libraries may be used to determine peptide ligands for opioid receptors. *NIDA Research Monograph 134*; National Institute on Drug Abuse: Washington, DC, 1993; pp 66−73.
7. Hruby, V. J.; Lam, K. S.; Lebl, M.; Kazmierski, W.; Hersh, E. M.; Salmon, S. E. Preparation of large peptide libraries with one peptide per bead and their use for the discovery of peptides that bind to acceptors. *NIDA Research Monograph 134*; National Institute on Drug Abuse: Washington, DC, 1993; pp 75−83.
8. Wilson, S. R.; Czarnik A. W., Eds. *Combinatorial Chemistry, Synthesis and Application*; Wiley: New York, 1997.
9. Ostresh, J. M.; Dorner, B.; Blondelle, S. E.; Houghton, R. A. Soluble combinatorial libraries of peptide, peptidomimetics, and organics: fundamental tool for basic research and drug discovery. *Combinatorial Chemistry, Synthesis and Application*; Wiley: New York, 1997; Chapter 11.
10. Purdie, N.; Province, D. W. Algorithms for the quantitative validation of chiral properties of peptides. *Chirality* **1999**, *11*, 546−53.
11. Purdie, N.; Province, D. W.; Johnson, E. A. Tripeptide discriminations using circular dichroism detection. *J. Pharm. Sci.* **1999**, *88*, 715−21.
12. Purdie, N.; Province, D. W.; Layloff, T. P.; Nasr, M. M. Algorithms for validating chiral properties of insulins. *Anal. Chem*. **1999**, *71*, 3341−46.
13. Purdie, N.; Province, D. W.; Johnson, E. A. A convenient assay method for the quality control of peptides and proteins. *J. Pharm. Sci.* **1999**, *88*, 1242−1248.
14. Wold, S.; Sjostrom, M. Chemometrics, present and future success, *Chemom. Intell. Lab. Syst.* **1998**, *44*, 3−14.
15. Towell, J. F., III; Manning, M. C. In *Analytical Applications of Circular Dichroism, Techniques and Instrumentation in Analytical Chemistry*; Purdie, N., Brittain, H. G., Eds.; Elsevier Scientific: Amsterdam, 1994; Vol. 14, pp 175−205.
16. Sigel, H.; Martin, R. B. Coordinating properties of the amide bond. Stability and structure of metal ion complexes of peptides and related ligands. *Chem. Rev*. **1982**, *82*, 384−426.
17. Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika* **1985**, *23*, 187−200.
18. Casy, A. F.; Parfitt, R. T. Enkephalins, endorphins, and other opiod peptides. In *Opiod Analgesics, Chemistry and Receptors*; Plenum: New York, 1986; Chapter 10, pp 333−384.
19. Schiller, P. W. Development of receptor-specific opioid analogues. *Progress in Medicinal Chemistry*; Ellis G. P., West, G. B., Eds.; Elsevier Science: Amsterdam, 1991; Vol. 28.

JS990210D