A Convenient Assay Method for the Quality Control of Peptides and Proteins

NEIL PURDIE,* DENNIS W. PROVINCE, AND ERIN A. JOHNSON

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

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

Abstract □ The development of a convenient and very accurate procedure with which to discriminate among subsets of structurally similar peptides and proteins, and measure enantiomeric purities with very good accuracy, has been described in a series of recent articles. A factor preventing its general application to all peptide forms is that comparisons were originally limited to closed subsets of structurally similar types, e.g., dipeptides, tripeptides, and insulin drug forms. In the most recent of these articles, a modification to the method was described which did enable the comparisons to be extended between sets, in particular the di-and tripeptides. That same modification is extended even further in this article to include additional di- and tripeptides, glycylglycine oligomers, insulin drug forms, and neuropeptides. The same principal component analysis treatment used for data reduction and statistical comparisons in prior work enables the discrimination among 49 of the total of 51 analytes investigated.

Introduction

It was originally demonstrated that members of a series of structurally analogous dipeptides[1] and a separate series of tripeptides[2] could be differentiated one from the other by (i) complexing pure forms of the analytes to Cu(II) in aqueous pH 13 solutions and (ii) measuring the visible range circular dichroism (CD) spectra. Complete differentiation within each series was not entirely possible on the basis of just the CD spectral data. It was only achieved when two novel data reduction algorithms were applied to the experimental data. The algorithms are peculiar only to CD data and are applicable in every circumstance where multiwavelength CD detection is used.

Differentiations within groups were based upon comparisons made among the individual members and how they differed from an arbitrarily chosen reference. For the di-and tripeptides these were glycylalanine (GA) and glycylglycylalanine (GGA), respectively. Differentiation between groups was not possible because there was no ligand common to both series. This was resolved by complexing the Cu(II) ion with the auxiliary ligand D-histidine which was made to undergo ligand exchange with the same peptide analytes under the same solution conditions.[2,3] The Cu(II)-D-histidine host complex in effect acts as a chiral derivatizing agent which produces a diastereoisomer, or mixed chiral complex, on partial ligand exchange with each peptide.

Reasons for choosing D-histidine are that it forms a very stable Cu(II) complex, it undergoes rapid ligand exchange, and the Cu(II)−D-histidine host complex has a fairly intense biphasic (i.e., positive and negative bands) visible CD spectrum which affords good analytical sensitivity and an extra element of selectivity to the analyses. Although less expensive, the L-histidine enantiomer has a CD spectrum for the Cu(II)-complex that bears a closer resemblance to the spectra for Cu(II)-L-peptide forms and therefore diminishes the discrimination capability.

A second reason for choosing an auxiliary ligand is that it, rather than the analytes, takes on the thermodynamic responsibility for solubilizing the Cu(II) in strongly basic (pH 13) solutions. For the same reason racemic tartrate is used in the classical biuret test[4] which, until recently, was a method of choice to measure total blood proteins.[5] Apart from the introduction of the CD detector, the only change between this test and the biuret procedure is replacing DL-tartrate with D-histidine. All of the accepted experimental procedures that pertain to the biuret test apply to this modification. The first metal to ligand attachment is made through the N-atom of the terminal amine, with subsequent chelation through binding of the amide nitrogens of the next two peptide bonds.[6] Protein aggregation in the presence of high pH or Cu(II) ion is not a problem.[4] Auxiliary ligand concentrations are always kept in large excess over the metal ion concentration.

An added analytical advantage to using a chiral auxiliary ligand is that the CD spectrum for the host is not baseline. As a result, analytical sensitivities are increased significantly allowing for the use of much smaller quantities of the analyte for each assay. With D-histidine as the auxiliary ligand, the quantity of the analyte material can be decreased by a factor of 10−100 relative to the amounts used earlier.[2,3] This is a significant savings factor if amounts of the analytes are very small to begin with as they invariably are in the production of trial quantities of new drug forms.

The strategy of using Cu−D-histidine as a chiral derivatizing agent as a way to enhance analytical selectivity was successfully demonstrated by the total discrimination among manufactured forms of human, porcine, human Lyspro, and bovine insulins and between the A- and B-chains of bovine insulin.[7] These constitute a unique set of compounds for evaluating the analytical selectivity (even specificity) of the method since the polypeptide sequence changes are minimal and as it turns out remote from the active site, which is the coordinating to the Cu(II) ion. Human and porcine insulins differ only in the identity of the B30, acid terminus, residues which are L-threonine and L-alanine, respectively. The sequence variation is as remote as it can possibly be from the B-amine terminus, the known binding site.[6] Human insulin and the human LysPro variant forms differ only by reversal of the B28−B29 lysine-proline sequence one place removed from the acid terminus

10.1021/js990209e CCC: $18.00
Published on Web 10/28/1999

of the B-chain. Bovine and human analogues differ by two residue substitutions at points A8 (alanine for threonine) and A10 (valine for isoleucine), positions that are part of one of the cyclical sequences in the molecules. Differentiation among these structural variants is routine.

From the obvious analytical specificity of the method toward the insulins,[7] it is evident that interactions other than just the primary coordination and chelation of the residues at the amine terminus are factors to be considered in donor−receptor interactions. A possible major influence might be long-range chiral−chiral interactions between host and analyte ligands. From an analytical perspective, the evidence suggests that the general selectivity among proteins will be very broad. The simplicity of the experimental procedure and the specificity of the data reduction and data handling algorithms make the method an attractive alternative to chromatography for QC applications by manufacturers and regulatory agencies.

The modified biuret reagent and detection procedure might ultimately qualify as an Analyte Specific Reagent (ASR) as it is defined by the Food and Drug Administration in the ASR Rule[8,9] and be applicable to peptides and protein forms.

A major, new frontier in the pharmaceutical industry is the focus on the therapeutic properties of peptide and protein drug forms.[10] Because the number of chiral centers has virtually no limit,[11] the magnitude of the chirality regulatory control problem is increased almost exponentially. Since chiral derivatizations will not produce a single diastereoisomer, even the very best chiral chromatographic methods face what are probably insurmountable challenges.[12] In these circumstances a bulk spectroscopic method has many advantages. The intent of this work was to get a better sense for how general the analytical selectivity is with a special focus on QC of peptides and proteins.

Experimental Section

Chemicals—The full complement of peptide and proteins analytes used for the study is (**1**) glycyl-L-alanine (GA), (**2**) glycyl-D-alanine (Ga), (**3**) glycyl-(L)-phenylalanine (GF) (**4**) glycylglycine (GG), (**5**) glycyl-(L)-histidine (GH), (**6**) glycyl-(L)-isoleucine (GI), (**7**) glycyl-(L)-lysine (GK), (**8**) glycyl-(L)-methionine (GM), (**9**) glycyl-(L)-proline (GP), (**10**) glycyl-(L)-tyrosine (GY), (**11**) (L)-alanyl glycine (AG), (**12**) (L)-alanyl-(L)-alanine (AA), (**13**) (L)-alanyl-(L)-tyrosine (AY), (**14**) (L)-tyrosylglycine (YG), (**15**) (L)-tyrosyl-(L)-alanine (YA), (**16**) (L)-tyrosyl-(L)-tyrosine (YY), (**17**) glycylglycyl-(L)-alanine (GGA), (**18**) glycylglycylglycine (GGG) (**19**) glycylglycyl-(L)-histidine (GGH), (**20**) glycylglycyl-(L)-isoleucine (GGI), (**21**) glycylglycyl-(L)-leucine (GGL), (**22**) glycylglycyl-(L)-phenylalanine (GGF), (**23**) glycyl-(L)-histidylglycine (GHG), (**24**) (L)-leucylglycylglycine (LGG), (**25**) (L)-tyrosylglycylglycine (YGG), (**26**) (GGGG), (**27**) (GGGGG), (**28**) (GGGGGG), (**29**) DSLET, (**30**) DTLET, (**31**) DADLE, (**32**) DAGO), (**33**) DALDA amide, (**34**) DPDPE amide, (**35**) CTAP amide, (**36**) DynorphinA (1−9), (**37**) DynorphinA (1−11), (**38**) DynorphinA (1−13), (**39**) DynorphinA (1−13) amide, (**40**) DynorphinB (1−13), (**41**) Met^5-enkephalin amide, (**42**) Leu^5-enkephalin, (**43**) Leu^5-enkephalin amide, (**44**) (D)-Ala^2-Leu^5-enkephalin amide, (**45**) β-endorphin, (**46**) ICI 174,864, (**47**) PLO 17 amide, (**48**) human insulin, (**49**) porcine insulin, (**50**) Lyspro human insulin, (**51**) bovine insulin.

The glycine oligomers and di- and tripeptides were Sigma Chemical Co. products. Peptides were reported to have an enantiomeric purity (EP) in excess of 99.8%. Neuropeptides **29**−**41** and **45**−**47** were provided by CHIRON. Leu^5-enkephalins were obtained from Sigma. Insulins were taken from manufactured lots from Lilly, Novo Nordisk, and Sigma. Reagent grade D-histidine was from Sigma Chemical Co. and reported to have an EP better than 99.8%. Reagent grade $CuSO_4 \cdot 5H_2O$ was obtained from Fisher Scientific.

Solution Preparations—Stock solutions of the reference Cu(II)−D-histidine complex in 0.10 M NaOH were prepared in which the Cu(II) and D-histidine concentrations were 20 mM and 80 mM, respectively. KI (30 mM) was added as a stabilizer.[4] Working solutions of copper-D-histidine were prepared by diluting aliquots from the stock by a factor of 10 with 0.10 M NaOH. To simplify any future quality control QC application by making it even more amenable to automation, we chose to use equal masses of the analytes rather than equal concentrations. Normalizing the concentration of the analyte to a single value is not a critical factor when the equilibrium constants for the ligand exchange processes differ as much as they might be expected to differ for the range of materials used in this study. Aliquots of 10.0 mg were added to the copper−D-histidine stock prior to its dilution with 0.10 M NaOH. From the range of molar masses 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. Throughout this range the signal to analyte concentration correlation is linear.[1−3] A mass as large as 10.0 mg is prohibitively high especially when the method is to be applied to trial drugs where quantities are very limited. The problem can be alleviated to a large degree by reducing the volume of the working solution. CD signals are relatively intense. With access to more modern CD instrumentation than the Model series used here, the cell volume could be reduced by at least 10−100 fold with no loss of signal quality.

CD spectra measured after ligand exchange are simple aggregates of weighted spectra for the host and mixed ligand complexes. Any unbound ligand does not absorb in the visible range. Formation constants and complexation stoichiometries are incidental to QC quantitative interpretations. The purpose of QC methods is to ensure that the presumed chemical and enantiomeric purities of commercial drug products meet the reference standard. Provided the specified procedure for the assay is accurately defined and followed in every detail, the purity will be determined by comparing the CD spectral data for a product lot with the spectrum for reference standard material.

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: 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.

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 nm and 682 nm. The SD values were 7.42 ± 0.07 mdeg and −214 ± 0.60 mdeg, respectively, within and between stocks.[2,7]

Results and Discussion

CD activity in the visible range for chiral Cu(II) complexes is a result of disymmetric perturbations of the ground and excited-state ligand field orbitals by the chiral ligands.[6] Three CD-active electronic transitions are reported to occur over the wavelength range of the visible absorbance band.[6] Bands in the UV range, attributable to only the chirality in the ligands, both bound and unbound, are typically broad, very intense, and quite insensitive to changes in the environment of the coordinating metal ion. The lack of selectivity is the major reason for not exploiting the obvious analytical sensitivity advantage that is inherent in the intense UV bands.

Cu(II)−D-Histidine and Peptide Complexes—The microsymmetry of the Cu(II) ion first coordination sphere is tetragonal, or more accurately, an axially distorted octahedron, distortion being a consequence of Jahn−Teller effects.[13] In strong base D-histidine is anionic and binds to Cu(II) via the terminal amine-N, the carboxylate functional group, and a pyrimidine N-atom, to form a tridentate 1:1 complex.[6] Other coordinate positions might be occupied by hydroxide ions. The stability of the complex and the favorable mass action maintained by keeping the ligand in large excess over the metal means that virtually all of

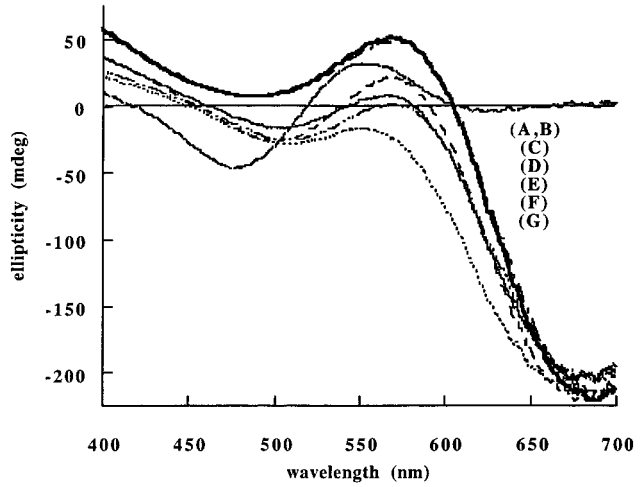

Figure 1—Visible range CD spectra for Cu(II) complexes with D-histidine and mixed D-histidine/dipeptides. (A) D-Histidine host, (B) GP, (C) GH, (D) GI, (E) GK, (F) GM, and (G) GF, arranged in order of the decreasing signal size at 570 nm. Curves A and B are superimposed and therefore indistinguishable.

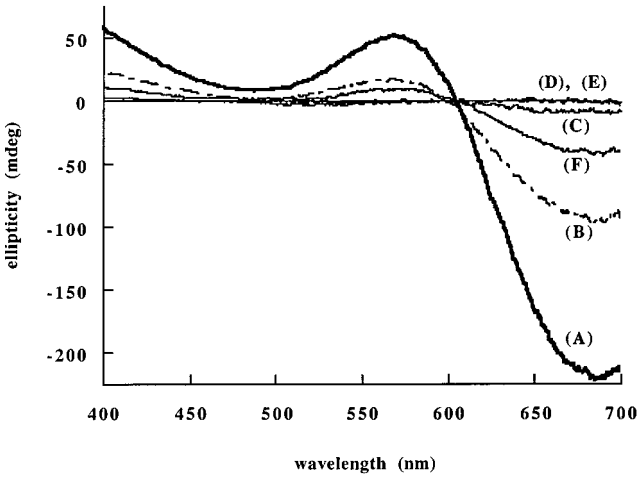

Figure 2—Visible range CD spectra for Cu(II) complexes with D-histidine and mixed D-histidine/oligoglycylglycines. (A) D-Histidine host, (B) GG, (C) GGG, (D) GGGG, (E) GGGGG, and (F) GGGGGG, arranged in order of the decreasing signal size at 700 nm.

the Cu^{2+} ion is in the form of the complex. The CD spectrum for the Cu(II)−D-histidine host chiral derivatizing agent is included for reference in Figures 1−5.

Complexing peptides to Cu(II) at high pH involves first attachment via the N-atom of the terminal amine group followed by ring closure(s) through bonding with the N-atoms of successive amide bonds until maximum thermodynamic stability is achieved.[6] Side-chain substituents on the amino acid residues lie out of the coordinate plane and are factors only in inter- and intramolecular interactions within the coordination sphere, unless a potential Lewis base is present, e.g., in L-histidine-containing peptides. Side-chain histidines are legitimate competitors for ligation positions, competing favorably with the terminal amine for the Cu(II) ion. The issue is moot if the test is restricted to QC where the binding mechanism would be the same for the reference standard and the sample. Metal to ligand stoichiometries for metal di- and tripeptide equilibria are 1:1 under pH > 12 conditions.[14] Coordinations with hydroxide ion and/or additional amide N-atoms in longer oligopeptides are variations that could affect the stoichiometry but not the analytical selectivity of the method.

Visible CD spectra for the Cu(II)−di and tripeptides in the absence of D-histidine were the objects of earlier work.[1,2]

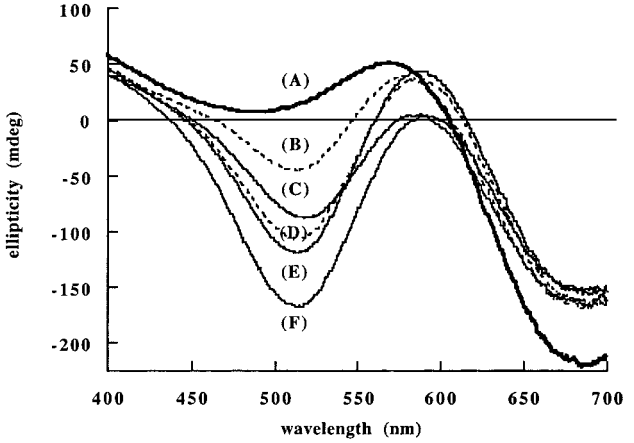

Figure 3—Visible range CD spectra for Cu(II) complexes with D-histidine and mixed D-histidine/Dynorphin neuropeptides. (A) D-Histidine, (B) Dynorphin B (1–13), (C) Dynorphin A (1–9), (D) Dynoprhin A (1–13), (E) Dynorphin A (1–11), and (F) Dynorphin A (1–13)-amide.

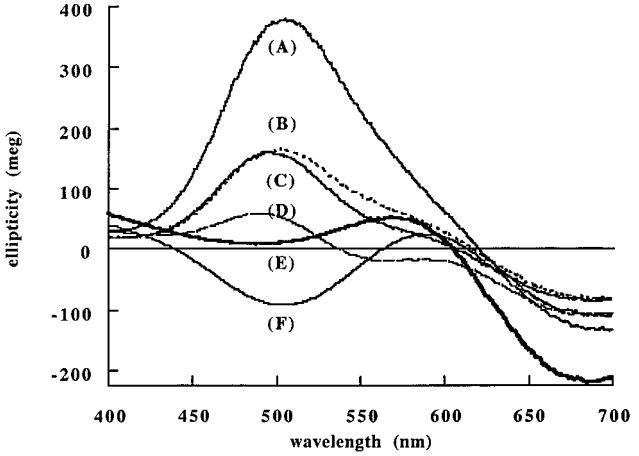

Figure 4—Visible range CD spectra for Cu(II) complexes with D-histidine and mixed D-histidine/enkephalin analogues. (A) a^2-Leu^5-enkephalin amide, (B) Leu^5-enkephalin amide, (C) Leu^5-enkephalin, (D) Met^5-enkephalin amide, (E) D-histidine, and (F) β-endorphin.

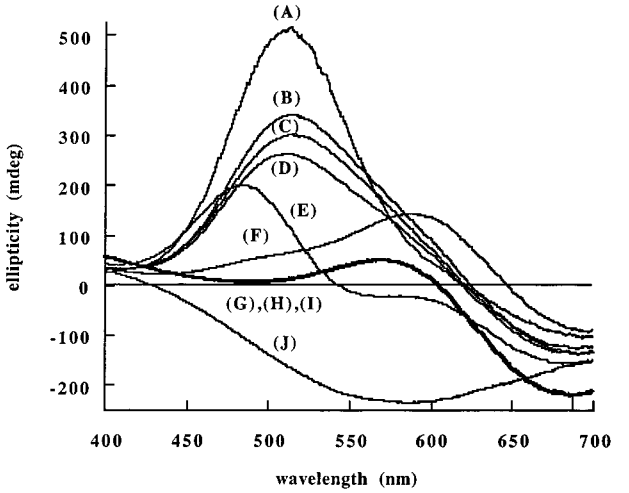

Figure 5—Visible range CD spectra for Cu(II) complexes with D-histidine and mixed D-histidine/modified neuropeptides. (A) DPDPE, (B) DTLET, (C) DADLE, (D) DSLET; (E) DALDA; (F) DAGO; (G) D-histidine, (H) ICI 174,864, (I) PLO 17, and (J) CTAP.

Spectra for Cu(II)−D-histidine and Cu(II)−(D-histidine-peptide) mixed complexes were the focus for di- and tripeptide ligands.[3,7] Evidence was provided that confirmed that the stoichiometries and the complex formation con-

stants were similar in magnitude within each series.[1,2] One would expect that for peptides to coordinate to the metal ion there has to be ligand exchange, or in some instances ligand addition with an expansion in the number of coordinating sites. Spectral variations that are the bases for the first level of analytical selectivity therefore are attributable to the individual stereochemical and conformational nature of the structures of the coordinated chiral ligands and to the relative rotational strengths of the three CD-active bands. That the CD-active electronic transitions are also sensitive to other factors besides the coordination geometry is reported for the insulins[7] in the Introduction section, where it is clear that minimal changes in peptide sequences that are remote from the binding site significantly alter the CD spectrum for the Cu(II)−D-histidine host complex.

Cu(II)−D-Histidine Ligand Exchange with Peptides:—Ligand exchange, as it is used in this context, is intended to cover all the mechanistic processes by which a peptide can become part of the Cu(II) complex entity. Total ligand exchange is not a possibility since mass action heavily favors the D-histidine ligand. Since D-histidine is bonded at three coordination sites, three other positions of the distorted octahedron are open to peptide addition reactions. Between these extremes are combinations of exchange/addition mechanisms, in which D-histidine is partially substituted, e.g., by loss of coordination to the carboxylate, and peptide is fractionally added. Nor should the possibility of outer-sphere ligand−ligand coordination be overlooked, e.g., in the insulin series.[7] The product of this ensemble of events is a compound mixture of "unreacted" histidine complex (always in excess) and a concentration distribution of peptide-containing complexes that clearly will vary with the structural properties and solution concentration of the analyte. It is the variability in the last in particular that enhances the analytical selectivity of the procedure. The method is not intended to determine the identity of any potential impurity, only to indicate that purity standards set by regulatory agencies have or have not been met.

Spectra for the mixed complexes are simple sums of the CD spectra for the components. Those that incorporate dipeptides that were not reported before[1] are shown in Figure 1. Spectra for the glycyloligomers, dynorphins, enkephalins, and structurally modified enkephalins are shown in Figures 2−5, respectively. Insulin spectra were previously published.[7] Selectivity distinctions are manifest by shifts in wavelength, sign, and signal intensity in the D-histidine band that maximizes at 550 nm. Besides the first-order chirality changes that occur within the first coordination sphere, there are others that are associated with reorganization of the tertiary structure in solution. Its disruption on ligand exchange could cause specific amino acid residues to impinge upon the 3-D architectural structure of the coordination sphere (outer-sphere complexation), adding more complexity to a mechanistic interpretation, but more diversity to the discriminations.

The broad absorbance band of Cu(II) complexes in aqueous media is known to consist of three subbands.[6] All are CD-active. Signs vary depending upon the nature of the ligands which, when taken altogether, form the basis for a broad analytical selectivity. The one obstacle to achieving total selectivity in Figure 1, for example, is separating the spectrum for GP from the spectrum for the D-histidine host complex. Whether ligands have exchanged is not known but is unlikely. For the other dipeptides in Figure 1 the general tendency is toward more negative spectra.

The only addition to the original mixed-tripeptide series[2] is the spectrum for GGG, Figure 2. As the length of the glycine-oligopeptides is increased, the characteristic shape of the host CD spectrum is retained as its intensity is diminished from GG through GGG, reaching what is essentially zero for the tetramer and the pentamer, after which the signal increases in intensity for the hexamer to reach a value that is about one-fourth of the original host complex. The hexamer result was reproducible and is not an experimental artifact. A possible explanation might be derived from an analogous result observed in a study of a series of dye-derivatized oligosaccharides where the visible range CD induction was attributed to incipient helicity (a single turn) when the oligomer length exceeded the dimension of the tetramer.[15] Whatever the mechanism, it is clear that the length of a peptide is a factor that is significant in the overall binding mechanism.

An expansion on the conclusions from the insulin study is manifested by the five Dynorphins all of which have the identical amino acid sequence through the first seven residues (YGGFLRR), far beyond what is believed to be the minimum for saturating the coordination positions on the Cu(II) ion. For the Dynorphin A analogues, the first *nine* residues in the sequence are identical, yet the spectra are easily distinguishable, Figure 3. The spectral response to an increase in length for the Dynorphin A neuropeptides is also nonmonotonic and most easily seen at 510 nm, the 1−11 analogue producing the greatest spectral change. There can be no question that ligand tertiary structures and their mutual molecular interactions with the primary architecture of the D-histidine complex are major factors in binding and therefore in broadening the analytical selectivity.

This last observation is an interesting concept to consider in the context of understanding factors that contribute to quantitative structure−activity relationships (QSAR). Interpretations of spectral changes should perhaps be revised to include, even emphasize, ligand−ligand interactions that do not involve their exchange. In other words, the *intact* Cu(II)−D-histidine complex might function as a model for a "receptor site" for peptides. If there is any validity to this concept, the possibility exists that the panorama of CD spectral changes observed in Figures 3−5 might eventually be capable of establishing in vitro correlations of molecular structure with therapeutic function. In that case visible range CD would have an important potential role as an insinuative ASR probe to better learn of the subtleties of chiral−chiral interactions in QSAR.

Data for the enkephalin analogues and β-endorphin are plotted in Figure 4. With the exception of a^2-Leu^5-enkephalin, all begin with the same YGG sequence, as do the Dynorphins. The primary coordination mechanism, therefore, will involve the same three Lewis bases, yet the spectra are unique, including an easy distinction between the acid and amide analogues of Leu^5-enkephalin. Spectra for the pentamers are dominated by positive bands over the short wavelength range. Sign inversion by these pentamers, relative to the 30-peptide protein, β-endorphin, and all of the Dynorphins, might be related to a restriction of outer sphere ligand ligand interactions. Signal enhancement is observed when D-alanine is injected into the sequence in a^2-Leu^5-enkephalin. It is surprising, but significant, that an enantiomeric switch did not cause the spectrum to be inverted in the D-histidine complex environment. This speaks to the relative importance of the retention of the chirality of the receptor in the binding mechanism.

The structurally engineered enkephalins and a few miscellaneous neuropeptides are the subjects of Figure 5. There is no evidence that ICI 174,864 and PLO 17 had sufficient influence on the D-histidine host complex to alter its spectrum. The terminal amine of ICI 174,864 is diallyl-

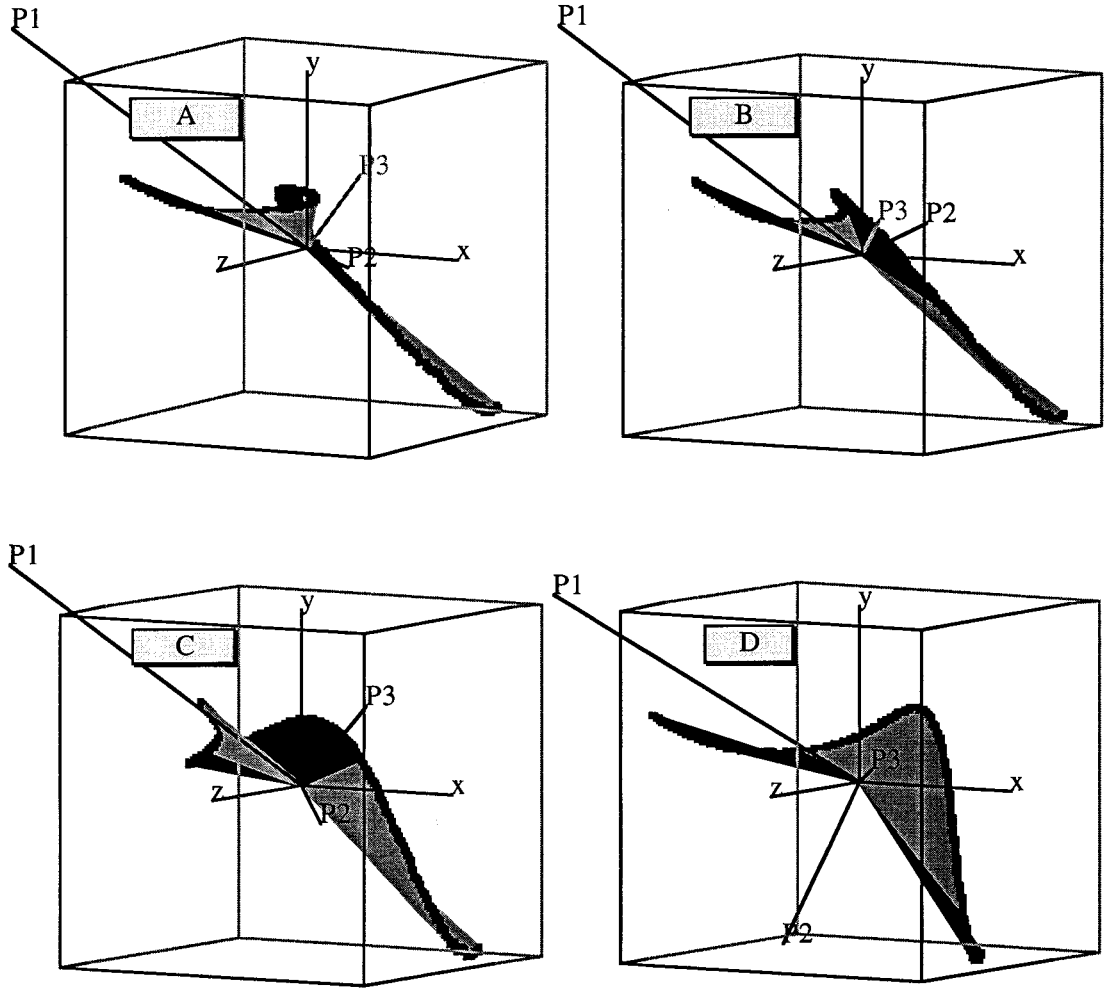

Figure 6—Representative plots resulting from the 3-D data reduction Spinning Plot algorithm for (A) Dynorphin A (1–13), (B) Dynorphin B (1–13), (C) a^2-Leu^5-enkephalin amide, and (D) CTAP. All four diagrams have the same *x*, *y*, *z* orientations. Dark lines that mark the perimeters of the curves mirror the CD spectra although the appearance is disguised by the orientation of the axes especially in A and B. P1, P2, and P3 are the eigenvectors for the principal components that are listed in Table 2. The distance from the origin to the extremity of each vector is the eigenvalue for that PC. Points to note are the different orientations in space for P2 ad P3 which are the most sensitive to the identity of the analyte.

substituted, and P^2 of PLO 17 is *N*-methylated. Both of these modifications adversely affect the primary coordination to the metal ion.

CTAP and DPDPE have disulfide rings and two D-enantiomers in their molecular structures. Between them they produce the greatest changes of any observed to date: strongly negative for CTAP and strongly positive for DPDPE. A positive signal for DPDPE is consistent with the effect of there being a D-enantiomer in the second position; see a^2-Leu^5-enkephalin. The N-terminal residue of CTAP is (D)F. There is insufficient information at this time to make a decision on whether the positional change from second to first is a major contributing factor in determining the sign of the CD spectrum.

The DSLET, DTLET, DADLE, DAGO, DALDA, and DPDPE modified enkephalins all have a D-amino acid residue in the second position. Similarities with the spectra for the natural enkephalins, Figure 4, exist, particularly the tendency toward strong positive bands at shorter wavelengths. Spectra for the DSLET, DTLET, and DADLE trio bear a strong resemblance to one another, but in what follows, they will be shown to be unique. For DALDA, a tetramer, and DAGO, which has an N-Me substituent on G3, structural differences are a bigger factor in determining spectral selectivity.

Data Reduction Algorithm—All the conclusions on analytical selectivity that have been reached so far in this paper are largely intuitive. For QC applications, a quantitative procedure is needed that will identify and/or verify a drug substance and measure its chemical and enantiomeric purities.

Two- and three-dimensional data reduction algorithms were introduced in prior work.[1,2,7] They are described as data reduction models because the objective was to reduce the 1500 data points, that make up the experimentally measured CD spectra, to a few definitive numbers which are characteristic of each analyte. Each procedure serves a special purpose. The 2-D model is effective in determining EP's with an accuracy unsurpassed by any other procedure.[1] The 3-D model, in contrast, is better suited for enantiomeric identification, and chemical purity determinations.[2,3] Only the former is considered in the context of this article.

The mathematical algorithm[16] used for the visual presentation of the 3-D data reduction is a three parameter Spinning Plot, available in a number of commercially available statistical software packages. The variables are wavelength (*x*-axis), CD data for the host complex (*y*-axis), and CD data for the mixed complexes (*z*-axis). Since CD signals are positive, negative, and zero, when two CD spectra are plotted, one against the other, four sign combinations are possible at any wavelength. Repeats of (*y*-, *z*-) coordinate points can occur at wavelength values, x_1 and x_2, that are not adjacent to each other in the spectra.

Table 1—Data Reduction by PCA of the Spinning Plot Data for the Histidine and (Histidine-a^2-Leu^5-enkephalin Amide) Mixed Complexes)

principal components	PC1	PC2	PC3
Eigenvalues	2.6942	0.2222	0.0836
Eigenvectors	–0.57655	0.59226	0.56287
	0.56405	0.78691	–0.25025
	0.59114	–0.17321	0.78776

When that occurs, plots are observed to "wrap-around" and become three-dimensional, Figure 6. To enhance the 3-D perspective, front and back quadrants are distinguished by dark and light shading. Distinctions are very evident.

With a total of 1500 data points for each variable, the system is overdetermined. In the data reduction procedure, data are subjected to a Principal Component Analysis (PCA) iterative process by which vectors of information are plotted against one another to derive principal components (PC's) which have length and directional properties that best describe the least-squares fit in the orthogonal directions of maximum variance. The number of PC's is equal to the number of variables, in this case three. PCA

Table 2—Eigenvalues from the PCA of Spinning Plots Data[a]

ligand	PC21	PC22	PC23	PC31	PC32	PC33
D-histidine	0.83992	0.38375	0.38375	0.00000	–0.70711	0.70711
AA	0.77506	0.61813	0.13116	0.29450	–0.53700	0.79051
AG	0.76020	0.64224	0.09811	0.32854	–0.51030	0.79477
AY	0.71550	0.69846	0.01430	0.40965	–0.42606	0.80127
YA	0.82951	0.25849	0.49507	0.14781	0.75322	–0.64094
YG	0.83631	0.29769	0.46060	0.10154	0.74112	–0.66365
YY	0.82003	0.51030	0.25913	0.15363	–0.63241	0.75925
Ga	0.80167	0.16171	0.57548	0.25970	0.77290	–0.57895
GA	0.83012	0.46494	0.30777	0.09642	–0.66336	0.74206
GG	0.83619	0.42950	0.34106	0.05456	–0.68392	0.72752
GY	0.80165	0.56357	0.19935	0.22146	–0.58973	0.77664
GH	0.00877	0.43018	0.90270	0.70517	0.63739	–0.31060
GK	0.81801	0.51753	0.25105	0.16291	–0.62704	0.76176
GI	0.83394	0.44463	0.32689	0.07247	–0.67544	0.73384
GP	0.84001	0.38129	0.38601	0.00293	0.70825	–0.70596
GM	0.82126	0.50619	0.26326	0.14870	–0.63535	0.75777
GF	0.75011	0.65664	0.07851	0.34779	–0.49267	0.79770
GGH	0.10485	–0.53847	0.83609	0.74200	0.60213	0.29475
GGI	0.02622	–0.63506	0.77202	0.73000	0.53978	0.41922
GGL	0.01357	–0.68830	0.72530	0.75108	0.48586	0.44702
GGF	–0.44698	–0.11093	0.88764	0.62046	0.67635	0.39696
GGA	–0.38508	–0.18286	0.90459	0.63138	0.66270	0.40274
LGG	0.31024	–0.02795	0.95025	0.67492	0.71043	–0.19945
YGG	0.41007	0.17842	0.89443	0.62412	0.66022	–0.41783
GHG	0.28915	0.26294	0.92046	0.64903	0.65295	–0.39040
GGG	0.79242	0.58094	0.18596	0.23681	–0.57395	0.78390
GGGG	–0.21947	0.18640	0.95765	0.70420	0.70962	0.02326
GGGGG	0.31758	–0.11740	0.94094	0.68218	0.71751	–0.14072
GGGGGG	0.83965	0.38788	0.38018	0.00477	–0.70522	0.70897
ICI 174,864	0.84001	0.38117	0.38613	0.00308	0.70831	–0.70590
PLO 17	0.83996	0.38273	0.38469	0.00121	0.70758	–0.70663
DynA (1–9)	0.82174	0.31009	0.47811	0.10080	0.74666	–0.65752
DynA (1–11)	0.64991	–0.07015	0.75677	0.50870	0.77995	–0.36457
DynA (1–13)	0.71161	0.00665	0.70254	0.42627	0.79079	–0.43926
DynA (1–13) amide	0.46401	–0.24232	0.85204	0.65364	0.74285	–0.14469
DynB (1–13)	0.83467	0.31635	0.45082	0.08315	0.73681	–0.67097
Met^5-enkephalinamide	0.59226	0.78691	–0.17321	0.56287	–0.25025	0.78776
Leu^5-enkephalin	0.58906	–0.18064	0.78764	0.56592	0.78798	–0.24253
Leu^5-enkephalinamide	0.68754	–0.03088	0.72549	0.45375	0.79829	–0.39603
a^2-Leu^5-enkephalinamide	0.60088	–0.13899	0.78716	0.55588	0.78031	–0.28656
β-endorphin	0.64855	–0.06949	0.75799	0.51106	0.77775	–0.36596
DSLET	0.72467	0.02654	0.68858	0.40629	0.79064	–0.45806
DTLET	0.68731	–0.02556	0.72591	0.46507	0.78316	–0.41276
DADLE	0.69716	–0.01355	0.71678	0.45019	0.78638	–0.42301
DAGO	0.70559	–0.00098	0.70862	0.47225	0.74621	–0.42301
DALDA	0.19164	0.79400	–0.57692	0.78289	0.23084	0.57776
DPDPE	0.51998	–0.20656	0.82883	0.62089	0.75778	–0.20067
CTAP	0.00507	–0.69886	0.71524	0.74454	0.48012	0.46384
human insulin	0.44975	0.83226	–0.32415	0.67287	–0.07705	0.73574
porcine insulin	0.42652	0.83526	–0.34701	–0.68620	–0.04888	0.72577
human Lyspro	0.31055	0.83468	–0.45483	0.73854	0.08936	0.66826
bovine insulin	0.55210	0.80567	–0.21466	0.59970	–0.20485	0.77356

[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 (Y.a. G(N-Me)FG); DALDA (YrFK amide); DPDPE (Ype*GFpe*); CTAP (fC*YwRTP*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); PLO 17 (YP(N-Me)Fp) amide.

solutions are expressed as three eigenvalues, over which the total variance is proportionated, and nine eigenvectors, that describe the 3-D spatial orientations of the eigenvalues. In the end, the original 1500 data points are reduced to three vectors of proportionately different lengths, the eigenvalues. The *x*- and *y*-variables are numerically the same for all analytes. Only the *z*-axis variable is analyte dependent. Taken as sets of three variables, however, the resultant PC's are all analyte dependent and as a result are the derived characteristic properties for standard reference materials against which all other preparations will be compared.

Eigenvalues and eigenvectors for PC1, PC2, and PC3, calculated for a^2-Leu^5-enkephalin amide, are given in Table 1. Their spatial projections are superimposed on the coordinate axes of Figure 6C. Eigenvectors give the most sensitive response to the identity of the analytes. Standard deviations in eigenvector values, calculated from repeated experimental measurements, are on the order of ±0.009, so considerable precision is attainable when making analytical differentiations or purity determinations.

Eigenvectors for all of the analytes are presented in Table 2. First on the list is the special case where spectral data for the host complex are plotted on both the *y*- and *z*-axes. This provides a common reference set against which values for all of the analytes can be compared. A cursory examination shows that the selectivity is least for PC11, PC12, and PC13, so the initial focus should be on the remaining six. Comparing only P22, P23, P31, and P32, the ambiguities are so few in number that it is claimed that 49 of the original 51 analytes are statistically and individually identifiable.

Nonlinearities observed in the spectral data for the glycyloligomer and Dynorphin A series are quantified in the results of the PCA.

The two exceptions to achieving total specificity among this group of polypeptides that range from 2 to 51 amino acid residues are ICI 174,846, and PLO 17. Differences between the factors for these analytes and the D-histidine reference are not statistically significant, implying that the host complex remains impervious to substitution. Sign inversions do occur for P32 and P33, but these are associated with the shortest eigenvalue PC3 and are open to uncertainty in interpretation. The same argument might be true for GP.

Quantitative applications of the 3-D Spinning Plot data reduction algorithm are discussed in prior work.[1−3] Results from these articles show that correlations between PC23 values and concentration are linear, making the analytical selectivity not only qualitative but also quantitative.

Summary

By introducing an auxiliary chiral ligand as a substitute for the achiral tartrate ion in the biuret test, a chirality-sensitive test has been created for a series of polypeptides whose sequences range from 2 to 51 residues. Sequence variations exist at both termini and internally (for the insulins[7]). In all but two instances the CD spectrum for the host is significantly altered.

D-histidine is but one of many potential auxiliary ligands. All but two of the current analytes could function in the role of host ligand. How close the already exceptional selectivity achieved by the current method can approach analyte specific reagent (ASR) status might ultimately depend on the identity of the auxiliary ligand. As the significance of the subtleties of the interligand chiral−chiral interactions become better understood, there is reason to expect that CD detection and multivariate modeling used together could contribute to a better understanding of QSAR models and predictions, receptor binding mechanisms, and drug design.

References and Notes

1. Purdie, N; Province, D. W. Algorithms for the quantitative validation of chiral properties of peptides *Chirality* **1999**, *11*, 546−53.
2. Purdie, N.; Province, D. W.; Johnson, E. A. Tripeptide Discriminations Using Circular Dichroism Detection. *J. Pharm. Sci.* **1999**, *88*, 715−21.
3. Purdie, N; Province, D. W. Peptide quality control test using ligand exchange on Cu(II) ion and circular dichroism detection *J. Pharm. Biomed. Anal.*, submitted.
4. Kingley, G. R. The direct biuret method for the determination of serum proteins as applied to photoelectric and visual colorimetry *J. Lab. Clin. Med.* **1942**, *27*, 840−6.
5. Tietz, N. W. *Fundamentals of Clinical Chemistry*, 3rd ed.; Saunders: Philadelphia, 1987.
6. 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.
7. Purdie, N.; Province, D. W.; Layloff, T. P.; Nasr, M. M. Algorithms for validating chiral properties of insulins. *Anal. Chem.* **1999**, *71*, 3341−46.
8. *Federal Register 62243*; Food and Drug Administration: Washington, DC, November 1998.
9. Weiss, S. M. Clinical new Update. Mol. Diagn. **1998**, *3* (1), 63−5.
10. 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.
11. 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.
12. Ward, T. J.; Armstrong, D. W. In *Chromatographic Chiral Separations*; Zief, M., Crane, J., Eds.; Chromatographic Science Series, Vol.40; Marcel Dekker: New York, 1988.
13. Cotton, F. A.; Wilkinson, G. *Advanced Inorganic Chemistry. A Comprehensive Text*, 4th ed.; Wiley-Interscience, New York, 1980.
14. Martell, A. E.; Sillen, L. G. *Stability Constants*; Special Publication No. 17; The Chemical Society: London, 1964.
15. Engle, A. E.; Hyatt, J. A.; Purdie, N. Induced Circular Dichroism Study of the Aqueous Solution Complexation of Cello-Oligosaccharides and Related Polysaccharides with Aromatic Dyes. *Carbohydr. Res.* **1994**, *181*, 265.
16. Kaiser, H. F. The varimax criterion for analytic rotation in analysis. *Psychometrika* **1985**, *23*, 187−200.

JS990209E