Multiple Conformations of the Proline-Rich Protein/Epigallocatechin Gallate Complex Determined by Time-Averaged Nuclear Overhauser Effects

Adrian J. Charlton,[†,§] Edwin Haslam,[‡] and Michael P. Williamson*,[†]

Contribution from the Department of Molecular Biology and Biotechnology, University of Sheffield, Firth Court, Western Bank, Sheffield S10 2TN, U.K., and Department of Chemistry, University of Sheffield, Western Bank, Sheffield S3 7HF, U.K.

Received December 3, 2001

Abstract: The structure of the complex between the heptapeptide Gln-Gly-Arg-Pro-Pro-Gln-Gly and the polyphenol (−)-epigallocatechin gallate (EGCG) has been determined using time-averaged nuclear Overhauser effects. Effective parameters for the force constant and time constant have been derived, allowing rapid and efficient calculation of structures that satisfy the input restraints. By using multiple start conformations, it is shown that conformational space is covered adequately and that the complex exists in one major conformation, in which the A ring of the EGCG is positioned over Pro5 and the D ring is over Pro4, with the B ring frequently close to the arginine side chain. Alternative conformations are also found, in which the prolines are almost always both involved in stacking interactions, with a strong preference for Pro4 to be involved. The structures are consistent with previous models for the interaction and suggest how precipitation of the complex could occur, which leads to the oral phenomenon of astringency. The method has promise as a general way of docking ligands onto receptors.

Introduction

Proline-rich peptides often adopt an extended type II polyproline helix[1] and play important roles in signaling pathways.[2] This is achieved through the proline residues, which act not only as hydrophobic surfaces but also as rigid links that reduce the loss in free energy on binding by keeping the free peptide relatively stiff.[1] A class of proline-rich peptides known as basic salivary proline-rich proteins is secreted by the human parotid salivary gland in large quantities, whose only established function is as a defense against dietary polyphenols, also known as tannins. Polyphenols have been divided into two major classes:[3] the hydrolyzable polyphenols, which are galloyl esters of glucose and are typically found in fruits, and the proanthocyanidins, which are found in a large variety of foods and beverages, including wine and tea, the latter being a rich source of (−)-epigallocatechin gallate (EGCG, Figure 1).[4] The interaction between proline-rich peptides and polyphenols has an important biological function: to reduce the bioavailability of higher molecular weight polyphenols and therefore reduce their toxic
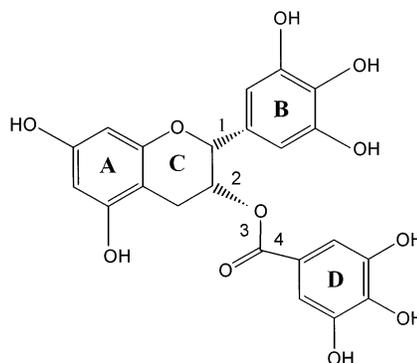
Figure 1. Structure of (−)-epigallocatechin gallate (EGCG). The four angles that define the major degrees of internal rotational freedom are indicated as angles 1−4.

effects in inhibiting digestive enzymes and complexing to iron.[5] The interaction in the mouth is thought to be responsible for the astringent taste of tea:[6] salivary proline-rich proteins bind to the polyphenols in the mouth and form insoluble complexes, which coat the surface of the mouth and form an astringent layer. In an effort to understand these phenomena at a molecular level, we have undertaken an NMR analysis of the complex between EGCG and a model proline-rich peptide, Gln-Gly-Arg-Pro-Pro-Gln-Gly, which represents a typical basic proline-rich protein repeat.

* Corresponding author. Phone: +44 114 222 4224. Fax: +44 114 272 8697. E-mail: m.williamson@sheffield.ac.uk.
† Department of Molecular Biology and Biotechnology.
‡ Department of Chemistry.
§ Current address: Central Science Laboratory, Sand Hutton, York YO41 1LZ, UK.

(1) Williamson, M. P. *Biochem. J.* **1994**, *297*, 249−260.
(2) Kay, B. K.; Williamson, M. P.; Sudol, M. *FASEB J.* **2000**, *14*, 231−241.
(3) Haslam, E. *Plant polyphenols: vegetable tannins revisited*; Cambridge University Press: Cambridge, UK, 1989.
(4) Charlton, A. J.; Baxter, N. J.; Haslam, E.; Williamson, M. P. In *COST 916: Bioactive plant cell wall components in nutrition and health—Polyphenols in Foods*; Amadò, R., Andersson, H., Bardócz, S., Serra, F., Eds.; European Commission: Luxembourg, 1997; pp 179−185.
(5) (a) Mehansho, H.; Butler, L. G.; Carlson, D. M. *Annu. Rev. Nutr.* **1987**, *7*, 423−440. (b) Salunkhe, D. K.; Chavan, J. K.; Kadam, S. S. *Dietary tannins: Consequences and remedies*; CRC Press: Boca Raton, FL, 1990; pp 113−146. (c) Bennick, A. *Crit. Rev. Oral Biol. Med.* **2002**, *13*, 184−196.
(6) Haslam, E.; Williamson, M. P.; Baxter, N. J.; Charlton, A. J. *Recent Adv. Phytochem.* **1999**, *33*, 289−318.

Even a cursory glance at the nuclear Overhauser effects (NOEs) observed in the complex shows that they are incompatible with any single structure. When the measured NOEs were used as structural restraints in a standard restrained molecular dynamics/simulated annealing calculation, the structures generated had numerous NOE restraint violations. However, more importantly, the structures of both peptide and polyphenol were forced into distorted and essentially meaningless conformations by the self-contradictory restraints.[7] In particular, the EGCG ring C was forced into an energetically unfavorable twisted boat conformation. If we are to derive structural information on the complex, we must therefore use some method that avoids the assumption of a single preferred conformation. Several methods have been adopted for analyzing multiple conformations.[8] Most of these methods consist of generating a large number of possible conformations, for example by molecular mechanics[9] or by a grid search,[10] and selecting a subset that are consistent with experimental results. Both of these structure generation protocols suffer from the problem that it is difficult to cover conformational space adequately except in cases of very few degrees of freedom:[11] either the description of a "conformation" remains vague (covering a large volume of conformational space),[12] or constrained minimizations are necessary to reduce the number of conformations used.[7–9,13] We have therefore explored the use of a very general method for searching conformational space, namely time-averaged NOEs.[14–16]

The basis of the time-averaged NOE method is very simple. In the conventional method, it is assumed that

$$I = kr^{-6}$$

where r is the internuclear distance and I is the measured NOE intensity. However, if the internuclear distance varies during the measurement of the NOE, then the relationship between r and I is more complicated.[17] As a reasonable approximation when intramolecular motion is rapid, we can write

$$I = k\langle r^{-3}\rangle^2$$

where the angled brackets indicate a time average. In the conventional restrained molecular dynamics approach to structure determination, the NOE defines a potential V_{dc} such that

$$V_{dc}(r) \begin{cases} = \frac{1}{2}K_{dc}(r - r^0)^2 & \text{if } r > r^0 \\ = 0 & \text{if } r \leq r^0 \end{cases}$$

with K_{dc} as a force constant which is set in a more or less arbitrary manner so as to produce the desired result, r as the instantaneous distance at any point during the calculation, and r^0 as the target distance measured in an NOE experiment. Because of the imprecise calibration of NOE intensities, r^0 is normally specified as a range rather than a single distance. In a time-averaged calculation, one could in principle merely replace r by $\bar{r}$, the time-averaged value of r. The problem with this is that the calculation would be markedly influenced by its history, and as the time of the calculation increased, it would become insensitive to recent movement. To counteract this problem, Torda et al. introduced a novel method by which the time-averaged distance is given an exponentially decaying memory characterized by a time constant τ:[14a]

$$V_{dc}(\bar{r}) \begin{cases} = \frac{1}{2}K_{dc}(\bar{r} - r^0)^2 & \text{if } \bar{r} > r^0 \\ = 0 & \text{if } \bar{r} \leq r^0 \end{cases}$$

with $\bar{r}(t)$ defined as

$$\bar{r}(t) = \left(\frac{1}{\tau}\int_0^t \exp(-t'/\tau)[r(t - t')]^{-3}\,dt'\right)^{-1/3}$$

This method has great potential. However, in the past it has suffered a number of problems,[16,18] principally that (a) the optimum values of K_{dc} and τ are not obvious and (b) the method appears not to be good at sampling conformational space adequately: either it gets stuck in a local minimum and is unable to cross local energy barriers within a reasonable time, or it allows the structure to overheat locally, which results in a small number of atoms developing very high velocity and "going off to infinity", and the calculation fails. Here we report the results of such a study, discuss methods for overcoming some of the limitations of time-averaged NOEs, and demonstrate that there is a single major conformer of the complex (in which ring A of EGCG is close to Pro5 and ring D is close to Pro4, with ring B generally close to the Arg side chain), but that other conformations are also present.

Experimental Section

All spectra were acquired in 90% H_2O/10% D_2O on a Bruker DRX-500 spectrometer at 3 °C. The solution contained the peptide at 4 mM and EGCG at 10 mM. It also contained a small amount of trimethylsilyl propionate (TSP), which was used as a chemical shift reference (0.00 ppm). The ^1H resonances of Gln-Gly-Arg-Pro-Pro-Gln-Gly and EGCG were assigned using standard 2D NMR methods, in particular using TOCSY and ROESY spectra with mixing times of 100 and 300 ms, respectively. The ROESY spectrum was also used to obtain NOEs for the complex. Spin-locking was achieved using a continuous field of 25 (TOCSY) or 2.2 kHz (ROESY). All spectra were acquired using the States-TPPI method[19] and processed using sine-squared bell window functions with zero-filling in the indirect dimension. Peaks were measured, assigned, and integrated using routines in FELIX (Accelrys Inc., San Diego, CA). Two-dimensional spectra of the peptide indicated

(7) Kessler, K.; Griesinger, C.; Lautz, J.; Müller, A.; van Gunsteren, W. F.; Berendsen, H. J. C. *J. Am. Chem. Soc.* **1988**, *110*, 3393−3396.
(8) (a) Kim, Y.; Prestegard, J. H. *Biochemistry* **1989**, *28*, 8792−8797. (b) Bonvin, A. M. J. J.; Boelens, R.; Kaptein, R. *J. Biol. NMR* **1994**, *4*, 143−149. (c) Blackledge, M. J.; Brüschweiler, R.; Griesinger, C.; Schmidt, J. M.; Xu, P.; Ernst, R. R. *Biochemistry* **1993**, *32*, 10960−10974.
(9) Landis, C.; Allured, V. S. *J. Am. Chem. Soc.* **1991**, *113*, 9493−9499.
(10) (a) Martin, J. T.; Norrby, P.-O.; Åkermark, B. *J. Org. Chem.* **1993**, *58*, 1400−1406. (b) Nikiforovich, G. V.; Prakash, O.; Gehrig, C. A.; Hruby, V. J. *J. Am. Chem. Soc.* **1993**, *115*, 3399−3406.
(11) Kozerski, L.; Krajewski, P.; Pupek, K.; Blackwell, P. G.; Williamson, M. P. *J. Chem. Soc., Perkin Trans. 2* **1997**, 1811−1818.
(12) Nikiforovich, G. V.; Vesterman, B. B.; Betins, J. *Biophys. Chem.* **1988**, *31*, 101−106.
(13) Cicero, D. O.; Barbato, G.; Bazzo, R. *J. Am. Chem. Soc.* **1995**, *117*, 1027−1033.
(14) (a) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1989**, *157*, 289−294. (b) Pearlman, D. A.; Kollman, P. A. *J. Mol. Biol.* **1991**, *220*, 457−479.
(15) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. *J. Mol. Biol.* **1990**, *214*, 223−235.
(16) Lepre, C. A.; Pearlman, D. A.; Futer, O.; Livingston, D. J.; Moore, J. M. *J. Biol. NMR* **1996**, *8*, 67−76.
(17) Neuhaus, D.; Williamson, M. P. *The nuclear Overhauser effect in structural and conformational analysis*, 2nd ed.; Wiley: New York, 2000; pp 171−174.
(18) Fennen, J.; Torda, A. E.; van Gunsteren, W. F. *J. Biomol. NMR* **1995**, *6*, 163−170.
(19) Marion, D.; Ikura, M.; Tschudin, R.; Bax, A. *J. Magn. Reson.* **1989**, *85*, 393−399.

that the peptide is almost entirely in the β-sheet region of the (φ,ψ) distribution and showed no measurable change in intraresidue or sequential rotating frame NOEs, or in *J*-couplings, on addition of EGCG. On the basis of previous conformational[20,21] and theoretical[1] studies, it is assumed that the peptide is largely in the polyproline II helix conformation, both free and bound.

Restrained molecular dynamics calculations were carried out using the program XPLOR.[22] A model of the covalent structure of EGCG was constructed using Insight (Accelrys Inc.), and an approximation to the three-dimensional conformation of the molecule was achieved by minimizing the conformational energy using the AMBER force field in Insight. The structure was saved from Insight as an XPLOR protein structure file (PSF), which was edited manually such that each atom had a unique name. Additional information about planar geometries (impropers) and dihedral angles was added to the PSF. To use the PSF file as input for XPLOR calculations, it was necessary to specify bond lengths, bond angles, and force constants for all bonds and angles by also creating an XPLOR parameter file for EGCG. The bonding parameters for EGCG were determined by comparison with common molecules such as amino acids and from the crystal structure of $(-)$-epicatechin.[23] A PSF file was also created for the peptide, as well as a parameter file, which was based on the file parallhdg.pro but modified slightly, such that ϕ angles of $-78°$ and ψ angles of $146°$ were imparted on the peptide as equilibrium backbone dihedral angles, using a stiff force constant of 500 kcal/(mol·rad^2), the same value as that used to restrain bond angles. This extended the peptide into a polyproline II helix,[1] in agreement with our previous results,[20] but allowed the backbone some flexibility and gave the side chains complete flexibility. Starting structures for calculations were generated by carrying out short molecular dynamics calculations with the EGCG in arbitrary positions in the absence of NOE restraints. In some runs, the EGCG structures did not have the correct chirality at one chiral center and were rejected as it was found that the chirality could not be corrected during further structure refinement. Twenty structures were chosen that did have the correct chirality and were used as input coordinate files for time-averaged NOE molecular dynamics simulations. The final calculations used a 300-step Powell minimization of the starting structure, followed by 25 ps of dynamics at a bath temperature of 300 K, using time steps of 0.5 fs. There then followed the time-averaged restrained NOE calculation, which continued over 40 ps using time steps of 1 fs, at a bath temperature of 300 K. All atom masses were set to 100 Da, and no solvent was present in the calculation. There is thus little relation between calculation time and real time: structures in the calculation can change conformation much more quickly than structures are known to do in solution at 300 K. Each 40-ps trajectory took approximately 30 min on a Silicon Graphics Indigo workstation.

The results of each time-averaged molecular dynamics simulation were recorded in a trajectory file, from which 50 PDB files containing snapshots of the peptide/EGCG complex were obtained using an XPLOR script. The separation of any two atoms and the bond or dihedral angle at any position in the complex could be extracted from the trajectory file using other XPLOR scripts. Structures were visualized using RasMol 2.6, and displayed by converting them to Raster3d images.

Results

Optimal Parameter Values. Fifty intermolecular NOEs were observed in ROESY spectra between the peptide and EGCG and were clearly incompatible with a single structure for the
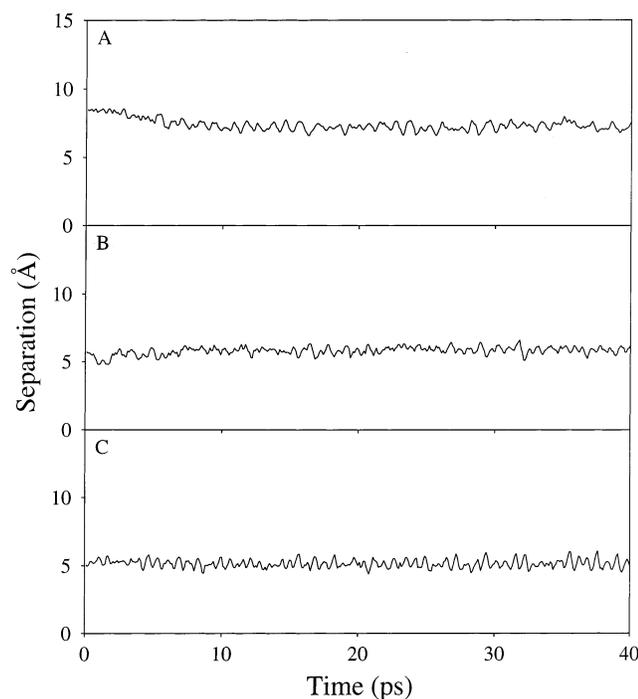
(20) Murray, N. J.; Williamson, M. P. *Eur. J. Biochem.* **1994**, *219*, 915–921.
(21) Baxter, N. J.; Lilley, T. H.; Haslam, E.; Williamson, M. P. *Biochemistry* **1997**, *36*, 5566–5577.
(22) Brünger, A. T. X-PLOR Version 3.1: A system for crystallography and NMR, Yale University, New Haven, CT, 1992.
(23) Fronczek, F. R.; Gannuch, G.; Mattice, W. L.; Tobiason, F. L.; Broeker, J. L.; Hemingway, R. W. *J. Chem. Soc., Perkin Trans. 2* **1984**, 1611–1616.

Table 1. List of NOEs between EGCG and the Proline-Rich Peptide[a]

peptide proton	EGCG proton
P4 CδH$^{1/2}$ or G7 CαH$^{1/2}$	2″/6″
P4 CδH$^{1/2}$ or G7 CαH$^{1/2}$	6/8
P4 CαH or R3 CαH	2″/6″
P4 CαH or R3 CαH	6/8
P4 CγH$^{1/2}$ or P5 CγH$^{1/2}$ or Q6 CβH$^{1/2}$	2″/6″
P4 CγH$^{1/2}$ or P5 CγH$^{1/2}$ or Q6 CβH$^{1/2}$	6/8
P4 CγH$^{1/2}$ or P5 CγH$^{1/2}$ or Q6 CβH$^{1/2}$	2
P4 CγH$^{1/2}$ or P5 CγH$^{1/2}$ or Q6 CβH$^{1/2}$	4α/4β
P4 CγH$^{1/2}$ or P5 CγH$^{1/2}$ or Q6 CβH$^{1/2}$	3
P4 CβH$^{1/2}$ or P5 CβH$^{1/2}$ or Q1 CβH$^{1/2}$	2″/6″
P4 CβH$^{1/2}$ or P5 CβH$^{1/2}$ or Q1 CβH$^{1/2}$	8/6
P4 CβH$^{1/2}$ or P5 CβH$^{1/2}$ or Q1 CβH$^{1/2}$	4α/4β
P4 CβH$^{1/2}$ or P5 CβH$^{1/2}$ or Q1 CβH$^{1/2}$	2′/6′
P4 CβH$^{1/2}$ or P5 CβH$^{1/2}$ or Q1 CβH$^{1/2}$	2
P5 CδH$^{1/2}$	2″/6″
P5 CδH$^{1/2}$	6/8
P5 CδH$^{1/2}$	2
R3 CδH$^{1/2}$	2″/6″
R3 CδH$^{1/2}$	2′/6′
R3 CδH$^{1/2}$	6/8
R3 CδH$^{1/2}$	3
R3 CβH$^{1/2}$ or R3 CgH$^{1/2}$	8/6
R3 CβH$^{1/2}$ or R3 CgH$^{1/2}$	2″/6″
R3 CβH$^{1/2}$ or R3 CgH$^{1/2}$	6/8
Q6 CγH$^{1/2}$	6/8
G2 CαH$^{1/2}$ or Q1 CαH	2″/6″
G2 CαH$^{1/2}$ or Q1 CαH	6/8

[a] All NOEs were specified using ambiguous restraints[24] and had an upper bound of 5 Å.

complex. For example, NOEs are observed between the Pro4 CγH protons and EGCG protons 6″/2″, 6, and 8, which cannot all occur in a single structure without excessive strain to the structure. Many of the NOEs are weak, and there is significant chemical shift overlap in the peptide spectrum, meaning that many of the NOEs had to be entered as ambiguous restraints.[24] In preliminary calculations, it was found to be difficult to achieve convergence of the calculations using conventional strong/medium/weak categories of NOE intensity, and therefore all restraints were entered with an upper distance bound of 5 Å. The restraints used are listed in Table 1.

It was found to be crucial to use an appropriate value for the force constant, K_{dc}. Too strong a value effectively removes the time dependence, and the calculation takes the appearance of a conventional NOE-restrained calculation. Figure 2 shows the effect that weighting the NOE term too strongly has on the separation between the Pro4 CαH and aromatic protons from the three rings of EGCG (Figure 1). The conformational space occupied by the EGCG protons at these three positions is severely restricted, and the advantages of averaging the NOEs over time are lost. This trajectory represents something very similar to that which would be expected during a conventional structure calculation. By plotting dihedral angles within EGCG as a function of simulation time, it is easily shown that the EGCG molecule has only very limited flexibility during the calculation, in agreement with the postulate of over-restricted motion. By contrast, too weak a value of K_{dc} allows large NOE violations that are incompatible with the experimental observations. In this study, we were particularly concerned with producing structures for which the input restraints were satisfied. Our successful strategy therefore used a value just large enough
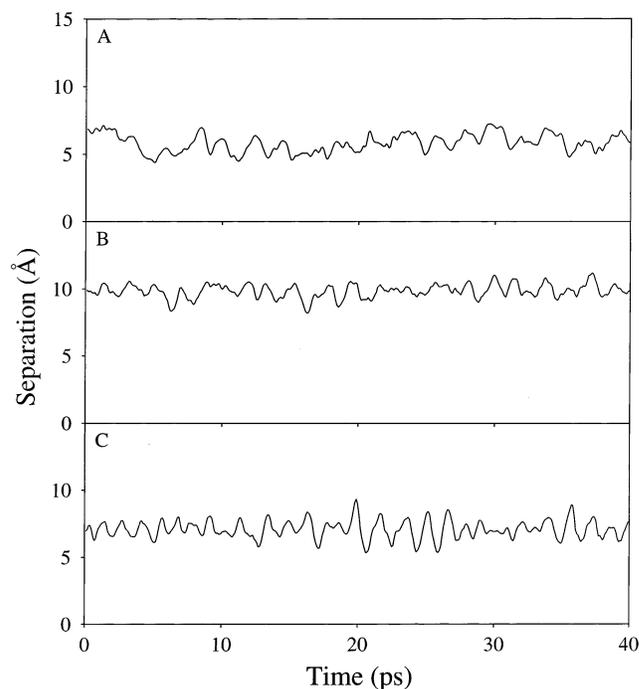
(24) Nilges, M. *J. Mol. Biol.* **1995**, *245*, 645–660.

Figure 2. Effect on the separation between Pro4 CαH and the C6H of (A) the D, (B) the B, and (C) the A rings from EGCG during the course of a time-averaged NOE-restrained molecular dynamics calculation, due to weighting the NOE term too strongly: the value of K_{dc} used was 500 kcal/(mol·Å^2), 10 times higher than the successful value, with τ set at 163 fs. Note that in this and subsequent figures, all structures have been aligned on the peptide backbone, which remains approximately fixed during the calculation.
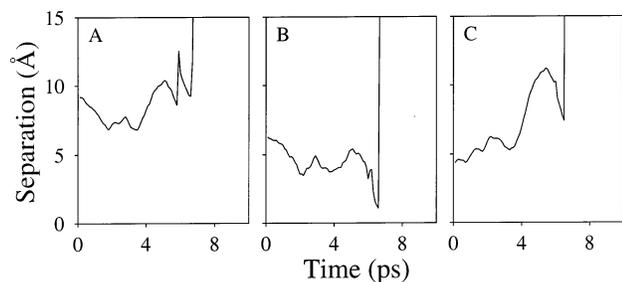
Figure 3. Effect on the separation between Pro4 CαH and the C6H of (A) the D, (B) the B, and (C) the A rings from EGCG during the course of a time-averaged NOE-restrained molecular dynamics calculation, due to setting τ too high [263 fs, compared to the successful value of 163 fs: K_{dc} was 50 kcal/(mol·Å^2)]. After approximately 7.5 ps of simulation, a number of atoms develop very high velocities, and the structure effectively falls apart.

to result in no NOE violations, namely 50 kcal/(mol·Å^2). It is worth noting, however, that at any individual time point, restraints may be violated by several angstroms.

Having set K_{dc}, it is then necessary to use an appropriate value of τ. Very short values of τ again allow no time-dependent variation of distance, while long values permit the structure such a large degree of conformational freedom that it rapidly overheats locally, and the calculation crashes. Because a large value of K_{dc} is used in the calculation, it is necessary to use a short value of τ to avoid overheating. Figure 3 shows the effect of using a value of τ that is too long: after approximately 7.5 ps of simulation, the structure falls apart. The best value of τ was found to depend on the number and upper bound of the NOE restraints and was determined by setting it to the highest

Figure 4. Separation between Pro4 CαH and the C6H of (A) the D, (B) the B, and (C) the A rings from EGCG during the course of a time-averaged NOE restrained molecular dynamics calculation performed with optimal values of τ and K_{dc} [$\tau = 0.163$ ps, $K_{dc} = 50$ kcal/(mol·Å^2)]. The reason the distance is oscillating around distances larger than 5 Å is that the distances are specified as ambiguous restraints: there is therefore no specific requirement for the time average of these individual distances to be less than 5 Å.

value possible that does not lead to the structure overheating (having already set a value for K_{dc}). For the final set of calculations reported here, it was 0.163 ps. This very short value was long enough to allow the EGCG molecule essentially complete internal rotational freedom, and restrained distances could vary widely without overheating (Figure 4). The inherent flexibility of the EGCG molecule during the trajectory is illustrated by Figure 5, which shows the four key dihedral angles that determine the structure of EGCG. Dihedral angles within EGCG covered essentially the entire range of ±180°. There is an approximately periodic oscillation apparent in the distances and angles, with a period of 1.5–2 ps, considerably longer than the time constant. A similar observation has been made in previous time-averaged calculations[15] and accounts for the recommendation (see below) that the trajectory should last at least 10 times longer than the time constant. Further analysis showed that there is no discernible correlation between the angles, again showing that EGCG has a large degree of internal rotational freedom. Although there were no explicit restraints on it other than those imposed by the covalent geometry, ring C of EGCG remained in the energetically preferred half-chair conformation throughout the calculation. This is presented as further evidence that there is no conformational strain imposed on the system by the NOE restraints.

Finally, it is necessary to use a long enough simulation time to cover all conformations readily accessible to the calculation. The length of time that was used for the simulation was critical only if a large decay constant was used. This was expected, as Torda et al.[15] reported that the length of the simulation should be at least 1 order of magnitude longer than τ. A simulation
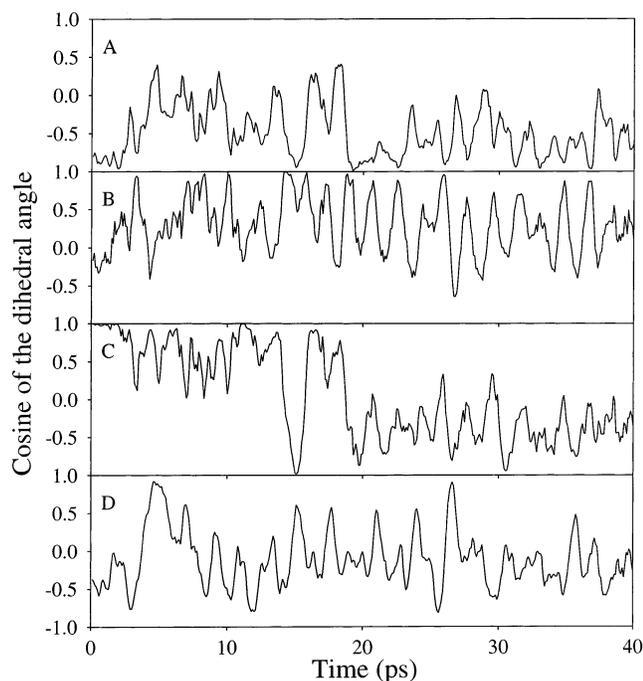
Figure 5. Time-averaged NOE trajectory showing the cosine of the four freely rotatable dihedral angles of EGCG when τ and K_{dc} are set to optimal values [$\tau = 0.163$ ps, $K_{dc} = 50$ kcal/(mol·Å^2)]. Panels A–D show the time dependence of angles 1–4 (Figure 1), respectively.

time of 40 ps was chosen, allowing each calculation to be performed reasonably quickly (about 30 min) without restricting the conformational space explored during the calculation.

Covering Conformational Space. Although Figure 4 suggests that there is relative translational freedom between the peptide and EGCG, an overlay of the resulting structures for a single trajectory (i.e., a trajectory starting from a single initial conformation) shows that the position of the EGCG is fairly localized (Figure 6A). This figure is already fairly crowded, making it difficult to interpret. To simplify the representation, the centers of mass for each of the three aromatic rings in EGCG have been calculated and plotted next to a representative peptide structure in Figure 6B. It is clear that the centers of mass *are* localized and that the D ring seems to stack over the Pro4 pyrrolidine ring, while the A ring falls over the Pro5 ring.

Over the course of the trajectory, all the time-averaged restraints were satisfied, with essentially no residual violations. For example, in the final structures, the contribution to the total energy from the NOE violations was less than 0.5%. Moreover, the time dependence of distances and angles shown in Figures 4 and 5 suggests that a stable set of states has been reached. Nonetheless, the fact that all of the structures sampled during the trajectory are so similar to each other suggests that conformational space may not have been adequately sampled. This result demonstrates one of the common problems identified in previous studies using time-averaged NOEs, that despite the time averaging, and the clear evidence of conformational freedom of the EGCG, the molecules are not able to search conformational space adequately. We therefore investigated whether other structures can be found that also satisfy the distance restraints, by calculating trajectories starting from different starting points. A total of 20 different starting points were used, with the EGCG in a random position and orientation, up to 20 Å away from the center of mass of the peptide. In all
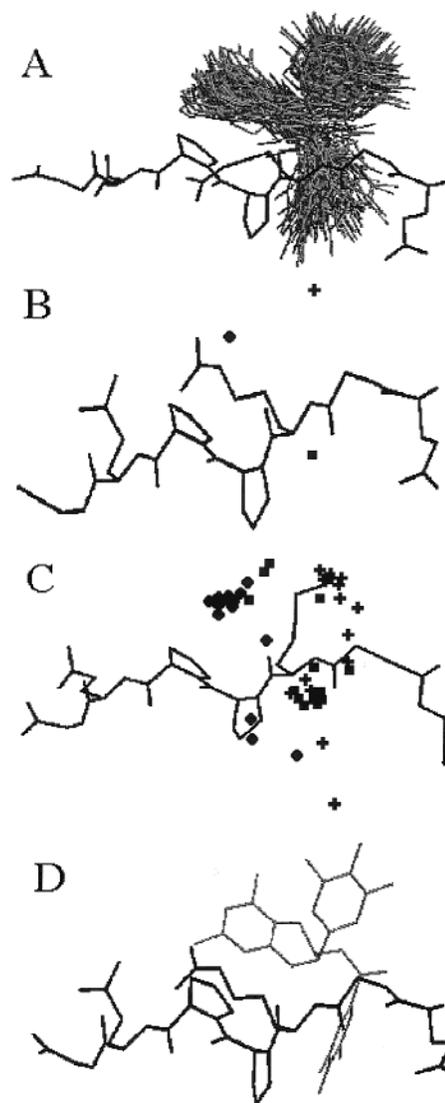
Figure 6. Steps used in the calculation of accessible structures. For each structure, a different peptide view is shown, to indicate that the peptide backbone remains relatively rigid, but the side chains are free to move. Note that the peptide is drawn from right to left, i.e., the *N*-terminal Gln is on the right. (A) An ensemble of 50 EGCG molecules, obtained as snapshots every 0.8 ps through the trajectory. (B) From this ensemble, the mean centers of mass of each ring were calculated. Rings A, B, and D are represented as circle, cross, and square, respectively. (C) The mean centers of mass of EGCG rings resulting from 20 independent trajectories. (D) A representative structure, being the structure nearest the center of the main cluster shown in (C).

20 cases, the complex rapidly reached a stable single family of conformations in which the time-averaged NOEs were satisfied. For each of the 20 trajectories, a mean position for each of the three aromatic rings of EGCG was calculated, as shown above for Figure 6B, and the resultant 20 × 3 ring centers are shown in Figure 6C. It is apparent that the ring centers remain in approximately the same position in many of the trajectories, since most squares are grouped together, as are most circles and most crosses. Therefore, it is reasonable to characterize the structure of a "typical" complex, which was calculated by the following procedure. Of the 20 mean structures, 14 are very similar to each other, in that the centers of each of the rings A, B, and D are in similar locations with respect to the peptide backbone. Therefore, a mean ring A position was calculated
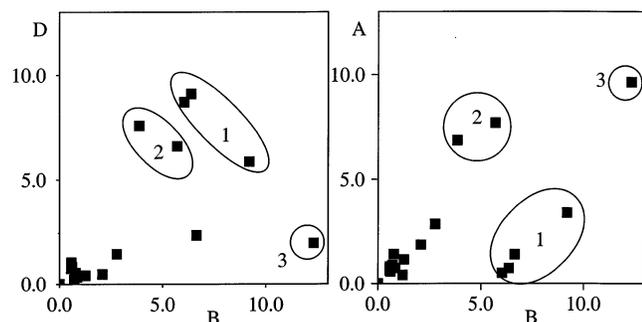
Figure 7. Two-dimensional representation of the distance (in angstroms) between the center of mass of each of the rings A, B, and D in Figure 6C and the corresponding ring in the representative "typical" structure (Figure 6D) (d_A, d_B, and d_D). The three groups showing unusual ring positions are marked 1, 2, and 3, and are discussed in the text.

from these 14 structures, as was a mean ring B position and a mean ring D position. For each of the 20 structures, we then calculated the average distance from the ring centers to these mean positions:

$$d = [d_A + d_B + d_D]/3$$

where for example d_A is the distance from the center of ring A to the mean position of ring A determined from the 14 similar structures. The structure with the smallest resultant value of d was picked as the "typical" complex, and the time point within this trajectory that was closest to the mean positions is thus a "typical" structure, and is shown in Figure 6D.

Six out of the 20 trajectories produced markedly different mean positions for the EGCG rings. To characterize the differences between trajectories, in Figure 7 we plot each structure in terms of the distance from the center of each aromatic ring A, B, and D to the center of the "typical" structure (d_A, d_B, and d_D as defined above). This means that in Figure 6, if a structure has all three rings close to the position of the rings in the typical structure, all three distances d_A, d_B, and d_D (denoted A, B, and D in Figure 7) will be short, and therefore the structure will be close to the origin in both plots. The six structures that are significantly far from the origin are ringed in Figure 7, and are discussed below.

Discussion

Strategy for Covering Conformational Space. Previous studies using time-averaged NOEs[25,26] have tended to stress the application of time averaging to reach a better understanding of molecular motion. They have therefore used rather weak force constants and long time constants, so as to perturb the dynamics as little as possible. By contrast, here we are not concerned with modeling dynamic processes, but only with producing a representative range of structures that are consistent with experimental data. The methodology used here to calculate the structures of the complex between the peptide Gln-Gly-Arg-Pro-Pro-Gln-Gly and EGCG can therefore be described as follows: (a) express all NOEs as weak distance restraints; (b) set K_{dc} as small as possible, *consistent with ending up with essentially zero time-averaged violations*; (c) set τ as long as possible, *consistent with the calculations not overheating locally*;

(d) repeat the calculations, using enough different trajectory starting points to cover conformational space adequately. The italicized phrases are crucial, because in order to obtain zero time-averaged violations, it is necessary to use a high value of K_{dc}, and in order to avoid local overheating, it is necessary to use a short value of τ. In particular, the optimal value of τ is 1−2 orders of magnitude smaller than those in most studies using time-averaged NOEs.[15,16,25,27] These values for K_{dc} and τ limit the already limited ability of time-averaged calculations to search effectively over conformational space within the time of the calculation, and the use of multiple starting positions is crucial to an effective search, as noted previously by other authors.[16,28] Clearly, the number of different trajectories required will depend on the complexity of the problem: in our case, the problem is relatively straightforward, because both the peptide and EGCG have a limited number of angles that can vary freely. (We estimate that the system can be approximated to 6 degrees of translational and rotational freedom between the two molecules, 4 rotational degrees of freedom within EGCG, and 5 rotational degrees of freedom of peptide side chain angles, giving 15 altogether.) The number of trajectories necessary in this study was therefore small. It is, of course, possible that other starting positions will result in different resultant structures, but the spread of results found in this study suggests that no radically new solutions will emerge.

The calculations reported here do not permit any inferences to be made about energy barriers and the rate of exchange between the conformers, because the time-averaged restraints are artificial. However, the NMR spectra show no sign of slow or intermediate exchange, and it is likely that exchange between the conformers is rapid.

The repeating of calculations has some similarities to the method of ensemble averaging[29] but has the distinct advantage that the result is not dependent on the choice of initial structures: all that matters is that the number of structures chosen should be large enough. Ensemble averaging generally requires weighting of the structures in the ensemble, for example according to their free energy or enthalpy. This can be done explicitly or via an additional adjustable parameter, and for this reason it has been suggested[18] that time averaging is a more elegant search strategy, provided that there is some way of efficiently searching around high energy barriers, such as the use of multiple start conformations. An alternative way of expressing the same idea is to note that, in ensemble averaging, the starting conformers form a closed set that constrain the combinations of allowed solutions, implying that altering the population of starting conformers can radically change the outcome. By contrast, in time-averaged calculations with multiple starts, inclusion of additional starting structures can only increase the range of acceptable solutions (provided that all solutions are free from violations, as observed here).

Finally, we note that the strategy adopted here is extremely rapid. Each trajectory takes only approximately 30 min on a small computer: the complete search used here therefore took only ca. 12 h. The conformational search was simplified greatly

(25) (a) Nanzer, A. P.; Poulsen, F. P.; van Gunsteren, W. F.; Torda, A. E. *Biochemistry* **1994**, *33*, 14503−14511. (b) Pearlman, D. A. *J. Biomol. NMR* **1994**, *4*, 1−16.
(26) Pearlman, D. A. *J. Biomol. NMR* **1994**, *4*, 279−299.

(27) González, C.; Stec, W.; Reynolds, M. A.; James, T. L. *Biochemistry* **1995**, *34*, 4969−4982.
(28) Worth, G. A.; Nardi, F.; Wade, R. C. *J. Phys. Chem. B* **1998**, *102*, 6260−6272.
(29) (a) Bonvin, A. M. J. J.; Boelens, R.; Kaptein, R. *J. Biol. NMR* **1994**, *4*, 143−149. (b) Cuniasse, P.; Raynal, I.; Yiotakis, A.; Dive, V. *J. Am. Chem. Soc.* **1997**, *119*, 5239−5248.

by using a conformationally restricted peptide backbone and allowing only the side chains and the EGCG to move. Calculations that allow the backbone more freedom are feasible but are much harder to analyze and therefore present severe difficulties in ensuring that conformational space is adequately covered. Although both we[21,30] and others[31] have found little or no evidence for any large-scale rearrangement of the peptide backbone on binding of small polyphenols, it may prove necessary to allow more backbone movement when studying complexes with larger polyphenols.

Analysis of Structures. In the "typical" structure, shown in Figure 6D, the A ring of the EGCG is positioned over Pro5, and the D ring is over Pro4. The B ring is in the vicinity of the arginine side chain; the exact distance depends, of course, on where the arginine side chain is, but it is close in many individual structures within the trajectory. The interaction between the EGCG ring and the proline is clearly a face-to-face stacking, with a hydrophobic character. This result is consistent with models that have previously been presented by us.[21,30,32] In these studies, chemical shift changes seen on titration of polyphenols into proline-rich peptides suggested that the major interaction is a stacking interaction between the phenolic ring and the proline side chain, with some interactions possible with arginine. A similar model was proposed on the basis of recent work with a wide range of amino acids.[33] Also in agreement with the findings presented here, our previous studies[21,33] suggested that polyphenols can bind in more than one location at the same time, which was suggested to strengthen the interaction between polyphenol and protein and to assist in the precipitation of polyphenol by salivary proline-rich proteins.

Although the majority of structures resemble the typical structure, three groups of "nontypical" structures were identified, as shown in Figure 7. Group 1, represented by three structures, has a small A distance but large B and D distances. Inspection of these structures shows that the A ring remains over Pro5,

but the B ring has moved into the position usually occupied by the D ring (i.e., close to Pro4). The D ring can occupy a range of positions. Group 2, represented by two structures, has all three distances large. Inspection of the structures shows that the D ring is now close to Pro5, and the A ring has moved into the cluster where typical structures have their D ring (close to Pro4); i.e., the A and D rings have interchanged compared to the typical arrangement. The B ring is forced into a different region by the swapping of the positions of rings A and D. Finally, group 3, represented by only one structure, has the D ring close to its position in the typical structure but both A and B rings displaced.

Thus, in 19 of the 20 sets of calculations, both prolines have stacking interactions with EGCG aromatic rings, these rings generally being A and D, with a preference for ring A to be stacked over Pro5 and ring D to be stacked over Pro4. The only interaction preserved in all 20 structures is a stacking interaction with Pro4. These results are again in agreement with our previous results, where face-to-face stacking centered on the first proline of a pair was deduced largely on the basis of chemical shift changes.[21] Interactions with the arginine side chain appear to be less important.

All the structures derived here present at least one sterically unhindered aromatic surface to the solvent. This is most often, but by no means always, ring B. In our previous studies[3,33] we have noted that precipitation of proline-rich protein/polyphenol complexes is likely to occur via interactions between exposed phenolic rings on one complex and proline or phenolic rings on the other. The present study offers a structural model by which this could occur.

In summary, the time-averaged calculations reveal a common face-to-face stacking mode of interaction between the two molecules but considerable variation in how this can be achieved in molecular terms. The methodology developed here is likely to be applicable to other multivalent docking interactions: it is quick and easy to implement and can be readily extended to incorporate other restraints.

(30) Murray, N. J.; Williamson, M. P.; Lilley, T. H.; Haslam, E. *Eur. J. Biochem.* **1994**, *219*, 923–935.
(31) (a) Hatano, T.; Hemingway, R. W. *J. Chem. Soc., Chem. Commun.* **1996**, *22*, 2537–2538. (b) Richard, T.; Verge, S.; Berke, B.; Vercauteren, J.; Monti, J. P. *J. Biomol. Struct. Dyn.* **2001**, *18*, 627–637.
(32) Charlton, A. J.; Baxter, N. J.; Lilley, T. H.; Haslam, E.; McDonald, C. J.; Williamson, M. P. *FEBS Lett.* **1996**, *382*, 289–292.
(33) Charlton, A. J.; Baxter, N. J.; Khan, M. L.; Moir, A. J. G.; Haslam, E.; Davies, A. P.; Williamson, M. P. *J. Agric. Food Chem.* **2002**, *50*, 1593–1601.