Carbopeptoid Folding: Effects of Stereochemistry, Chain Length, and Solvent**

Riccardo Baron, Dirk Bakowies, and Wilfred F. van Gunsteren*

The folding of a polypeptide chain into the stable three-dimensional structure of a biologically active protein is still not understood in atomic detail. However, several research groups have recently reported successful atomistic simulations of secondary-structure formation, including the formation of helices of different types, β turns and β sheets of α- and β-peptides.[1–23] Insight into the nature of both the folding process[24,25] and the unfolded state[15,26,27] has been obtained from various studies simulating the reversible folding of peptides. This development is encouraging and indicates that the biomolecular force fields in use are approaching the accuracy required to predict folding equilibria, although this has so far been demonstrated only for short polypeptides.

Experimentally, significant progress has been made in the design and synthesis of peptide analogues that mimic secondary-structure elements of proteins, such as α helices, turns, and β sheets.[28–30] For example, carbopeptoids, homo-oligomers of sugar-containing amino acids, have been prepared with both furanose[31] and pyranose[32] residues. These carbopeptoids are members of the family of δ-peptides which may formally be constructed from α-peptides by replacement of every second peptide fragment with a substituted tetrahydrofuran (THF) or tetrahydropyran ring.[33] These molecules have potential applications as drugs that block protein–protein interactions and inhibit enzyme catalysis.[34–36]

Structural preferences have been investigated in nuclear magnetic resonance (NMR) experiments.[31,37–40] Various oligomers with cis configurations at the C2 and C5 atoms of the THF ring (Scheme 1) tend to form conformations reminiscent of a conventional β turn.[41] The characteristic NH(i)/C0(i)/C0(i+2) hydrogen-bonding pattern has been observed for tetramer 1 in both chloroform and dimethylsulfoxide (DMSO).[37] Chain extension to six or eight residues does not alter the preferred secondary structure. Apparently, hexamer 2 and octamer 3 show the hydrogen-bonding pattern of tetramer 1, extended to six and eight residues, respectively. The trans-linked tetraters appear to have no conformational

[*] R. Baron, Dr. D. Bakowies, Prof. Dr. W. F. van Gunsteren
Laboratorium für Physikalische Chemie, ETH Zürich
8093 Zürich (Switzerland)
Fax: (+41) 1-632-1039
E-mail: wfvgn@igc.phys.chem.ethz.ch

[**] Financial support from the Schweizerischer Nationalfonds (Project no.: 2000-063590) and from the National Center of Competence in Research (NCCR), Structural Biology, of the Swiss National Science Foundation (SNSF) is gratefully acknowledged. We thank Dr. Tim Claridge for providing us with NMR spectroscopic data for the cis-tetramer 1.

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.
preferences or to adopt other types of structure, depending on the substituents at the C3 and C4 positions. The trans octamer 5 with ketal protecting groups at the C3 and C4 positions is reported to form a rare type of left-handed helix, stabilized by interresidue NH(i)–O(i−3) hydrogen bonds.

The apparent dependence of structural motifs on chain length and stereochemistry prompted us to carry out molecular dynamics (MD) simulations to supplement the indirect experimental observations with a more detailed atomistic picture. Scheme 1 shows the peptides considered in the present study. Tetramer 1, hexamer 2, and octamer 3 are all based upon a cis-linked β3-arabinofuranose scaffold. Hexamer 4 and octamer 5 are the corresponding trans-linked stereoisomers. All simulations were performed with chloroform as solvent. An additional simulation of tetramer 1 in DMSO was performed to assess the effect of solvent on secondary-structure formation.

The accuracy of MD simulations may be assessed by comparison to results from nuclear Overhauser effect (NOE) experiments. Qualitative NOE intensities have been reported for tetramer 1 in chloroform. Table 1 lists these data and compares them with

<table>
<thead>
<tr>
<th>Proton pair</th>
<th>NOE intensity</th>
<th>Model structure</th>
<th>(Trajectory)</th>
<th>(1st cluster)</th>
<th>(2nd cluster)</th>
<th>(3rd cluster)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N(4)–C6(3)-pro-R</td>
<td>w</td>
<td>0.325</td>
<td>0.422</td>
<td>0.432</td>
<td>0.423</td>
<td>0.361</td>
</tr>
<tr>
<td>N(4)–C5(4)</td>
<td>m</td>
<td>0.196</td>
<td>0.285</td>
<td>0.292</td>
<td>0.276</td>
<td>0.271</td>
</tr>
<tr>
<td>N(4)–C2(3)</td>
<td>w</td>
<td>0.352</td>
<td>0.430</td>
<td>0.316</td>
<td>0.325</td>
<td>0.336</td>
</tr>
<tr>
<td>N(4)–C4(4)</td>
<td>w</td>
<td>0.418</td>
<td>0.428</td>
<td>0.363</td>
<td>0.422</td>
<td>0.391</td>
</tr>
<tr>
<td>N(4)–C3(2)</td>
<td>w</td>
<td>0.441</td>
<td>0.276</td>
<td>0.280</td>
<td>0.248</td>
<td>0.241</td>
</tr>
<tr>
<td>N(3)–C6(3)-pro-S</td>
<td>s</td>
<td>0.227</td>
<td>0.262</td>
<td>0.280</td>
<td>0.248</td>
<td>0.241</td>
</tr>
<tr>
<td>N(3)–C6(3)-pro-R</td>
<td>m</td>
<td>0.285</td>
<td>0.267</td>
<td>0.252</td>
<td>0.281</td>
<td>0.284</td>
</tr>
<tr>
<td>N(3)–C6(2)-pro-R</td>
<td>m</td>
<td>0.280</td>
<td>0.400</td>
<td>0.355</td>
<td>0.408</td>
<td>0.407</td>
</tr>
<tr>
<td>N(3)–C5(3)</td>
<td>m</td>
<td>0.271</td>
<td>0.285</td>
<td>0.287</td>
<td>0.287</td>
<td>0.263</td>
</tr>
<tr>
<td>N(3)–C2(2)</td>
<td>w</td>
<td>0.340</td>
<td>0.327</td>
<td>0.327</td>
<td>0.327</td>
<td>0.327</td>
</tr>
<tr>
<td>N(3)–C4(3)</td>
<td>w</td>
<td>0.428</td>
<td>0.372</td>
<td>0.430</td>
<td>0.257</td>
<td>0.437</td>
</tr>
<tr>
<td>N(3)–C3(1)</td>
<td>w</td>
<td>0.427</td>
<td>0.654</td>
<td>0.507</td>
<td>0.687</td>
<td>0.715</td>
</tr>
<tr>
<td>N(2)–C6(2)-pro-S</td>
<td>s</td>
<td>0.230</td>
<td>0.253</td>
<td>0.240</td>
<td>0.265</td>
<td>0.257</td>
</tr>
<tr>
<td>N(2)–C5(1)</td>
<td>w</td>
<td>0.435</td>
<td>0.430</td>
<td>0.432</td>
<td>0.430</td>
<td>0.428</td>
</tr>
<tr>
<td>N(2)–C2(1)</td>
<td>w</td>
<td>0.328</td>
<td>0.321</td>
<td>0.328</td>
<td>0.317</td>
<td>0.319</td>
</tr>
<tr>
<td>N(2)–C4(2)</td>
<td>w</td>
<td>0.417</td>
<td>0.395</td>
<td>0.433</td>
<td>0.406</td>
<td>0.370</td>
</tr>
<tr>
<td>C4(3)–C6(3)-pro-R</td>
<td>s</td>
<td>0.250</td>
<td>0.283</td>
<td>0.249</td>
<td>0.349</td>
<td>0.256</td>
</tr>
<tr>
<td>C6(3)-pro-S–C5(1)</td>
<td>w</td>
<td>0.407</td>
<td>0.830</td>
<td>0.703</td>
<td>0.854</td>
<td>0.868</td>
</tr>
<tr>
<td>C6(3)-pro-S–C5(5)</td>
<td>s</td>
<td>0.245</td>
<td>0.253</td>
<td>0.236</td>
<td>0.285</td>
<td>0.241</td>
</tr>
</tbody>
</table>

[a] The first column shows the pairs of protons for which experimentally determined NOE data are available. Residue sequence numbers are given in parentheses (see Scheme 1 for reference). The NOE intensities in the second column are only classified as weak (w), medium (m), or strong (s).
sional structure, as they reflect an averaging over geometries biased \((r^{-3})^{-1/6}\) towards short proton–proton distances \(r\). Clearly this problem is more severe for peptides that show a high degree of conformational flexibility.

Figure 1 shows, as a function of time, the atom-positional root-mean-square deviation (RMSD) of trajectory structures of the 22-membered hydrogen-bonded ring. For the octamer, we show the formation of hydrogen bonds \(\text{NH}(4)\) cis to define the folded structures. For the tetramer, Dashed lines indicate the backbone RMSD similarity criterion chosen to build a model structure for the cis tetramer. The cis octamer (f) is compared to a model structure inferred from ref. [31]. The folded cis hexamer (e) is compared to hypothetical model structures derived from qualitative NOE data for the cis octamer. The trans octamer (f) is compared to a model structure inferred from ref. [31]. Dashed lines indicate the backbone RMSD similarity criterion chosen to define the folded structures. For the cis octamer (e), the formation of \(\text{NH}(3)\)–\(\text{O}(1)\) (red), \(\text{NH}(4)\)–\(\text{O}(2)\) (green), \(\text{NH}(5)\)–\(\text{O}(3)\) (blue), \(\text{NH}(6)\)–\(\text{O}(4)\) (black), \(\text{NH}(7)\)–\(\text{O}(5)\) (orange), and \(\text{NH}(8)\)–\(\text{O}(6)\) (cyan) hydrogen bonds is shown as a function of time. For the trans octamer (f), we show the formation of hydrogen bonds \(\text{NH}(4)\)–\(\text{O}(6)\) (red) and \(\text{NH}(7)\)–\(\text{O}(3)\) (green), which characterize, respectively, a 20- and 22-membered hydrogen-bonded ring.

from idealized model structures (see the Experimental Section). We have used all backbone atoms (N, C6, C5, O, C2, C) for this analysis, but excluded the C6, C5, and O atoms of the first residue and the C atom of the last residue (compare with Scheme 1). Single trajectory structures are considered folded if their RMSD from the reference model structure is less than 0.13 (tetramer), 0.17 (hexamers), or 0.28 nm (octamers).

The cis tetramer 1 experiences several folding and unfolding events during 100 ns of simulation. With chloroform as solvent, the molecule frequently adopts conformational close to the repeating β-turn right-handed helical model structure (Figure 1a). This conformation is less often visited when the solvent is DMSO (Figure 1b). The stronger hydrogen-bonding character of DMSO clearly disfavors formation of intramolecular hydrogen bonds in the peptide. While the hydrogen bonds \(\text{NH}(3)\)–\(\text{O}(1)\) and \(\text{NH}(4)\)–\(\text{O}(2)\) are, respectively, formed for 11 and 6% of the simulation with chloroform, they occur for only 1 and 3% of the simulation with DMSO as solvent (see the Supporting Information). Analysis of temperature coefficients and chemical shifts leads to similar conclusions.\([37]\)

The folded cis hexamer 2 forms four hydrogen bonds of type \(\text{NH}(i)\)–\(\text{O}(i-2)\) with \(i = 3, 4, 5,\) and 6. Figure 1c shows the RMSD time series calculated for our MD simulation with chloroform as the solvent. Several folding and unfolding events are observed after about 35 ns. We have not attempted to build a model structure for the trans hexamer 4, as no experimental data have been reported. Figure 1d shows that it assumes a conformation similar (RMSD < 0.17 nm) to the model of the cis isomer only once during the 100 ns of simulation. The trans substitution at the THF ring clearly disfavors the formation of the β turn found for the series of cis-linked carbopeptoids. Clustering trajectory snapshots into batches of highly similar configurations (see the Experimental Section) generally shows the dominant configurations sampled in an MD simulation. In the case of the cis hexamer 3, we find 11% of all snapshots in the first cluster and a total of 20 significantly (> 1% each) populated clusters, which sums up to 76% of all configurations. These results indicate that the ensemble of cis-hexamer configurations is dominated by a fairly small number of different types of structure. The trajectory generated for the trans hexamer, however, clusters into significantly more batches. Only 5% of all snapshots are found in the first cluster and a total of 30 significantly populated clusters cover only 62% of all configurations. We conclude that the trans hexamer shows a much lower tendency to form any preferred type of secondary structure than the cis hexamer. This observation is in line with the experimentally claimed\([49]\) absence of stable secondary structures for short trans-linked peptides of the type studied here.

The cis octamer 3 folds to a stable structure that shows the same hydrogen-bonding pattern, \(\text{NH}(i)\)–\(\text{O}(i-2)\), as the smaller cis oligomers. These hydrogen bonds are observed in our 100-ns-long simulation with chloroform as solvent (Figure 1e and the Supporting Information). Again, the trans octamer 5 shows very different behavior. \(\text{NH}(i)\)–\(\text{O}(i-2)\) hydrogen bonds are not formed at all, a result pointing to the steric constraints imposed by the different stereochemistry. However, other hydrogen bonds, \(\text{NH}(i)\)–\(\text{O}(i-k)\), occur with significant frequency (Figure 1f and the Supporting Information) to form rings of size 20 \((k = 2), 22 \,(k = -4), 26 \,(k = 3), 28 \,(k = -5), 32 \,(k = 4),\) and 38 \((k = 5)\). Reflecting their size, these rings are rather flexible and give rise to fairly large RMSD variations (Figure 1f). A left-handed helix composed of 16-membered rings \((k = -3)\) was suggested in experimental NMR studies for a similar trans octamer\([40]\), but this is
observed only once in our simulation (at 69 ns, Figure 1 f). Unfortunately, the experimental paper provides only vague information about the measured NOE values, thereby precluding a direct comparison between NOE and MD data. We note, however, that our analysis of the MD trajectory does not give any indication of long-range contacts of type NH(i)–H5(i–2), NH(i)–H3(i–3), H2(i)–H4(i–2), or H2(i)–H3(i–2), mentioned in ref. [40] as the calculated values of \((r^{-3})^{-1/2}\) range from 0.7 to 1.2 nm (data not shown).

Figure 2 shows the dominant conformations of the cis tetramer 1 and of both the octamers 3 and 5, together with the corresponding model structures. For the cis tetramer, there are 11 clusters populated by at least 1%, which totals 94% of the entire ensemble. The helical model structure falls into the third cluster, whose central member structure shows a similar backbone configuration. For the cis and trans octamers, we find a larger number of significantly (> 1%) populated clusters (27 and 28, respectively), which totals 82 and 87% of the entire ensemble, respectively. The larger diversity of structures reflects the increased conformational flexibility of longer peptide chains. In the case of cis octamer 3, the central member structure of the third cluster displays three examples of the characteristic NH(i)–O(i–2) hydrogen bonds \((i = 3, 4, \text{and } 6)\) discussed above. One hydrogen bond of this type is also present in the central member structures of the first and second clusters \((i = 6 \text{ and } 3, \text{ respectively})\).

Carbopeptoids containing THF moieties in their backbone show distinct folding behavior depending on chain length and absolute configuration at the THF–peptide link. The simulations presented herein show good agreement with the, admittedly limited, experimental data available. Sets of NOE data were published only for the tetramer, and they are in good agreement with the simulation. More qualitative conclusions drawn in experimental work are confirmed in the simulations as well. It is important to note that we have not undertaken any attempt to calibrate the force-field parameters specifically for carbopeptoid simulations. Instead we have used a well-calibrated general biomolecular force field which apparently captures the folding behavior of carbopeptoids. Once the agreement with experimental data is established, MD simulations of peptides can be used with some confidence to interpret experimental results in terms of conformational distributions and to explore the various types of secondary structure involved.

Two other major conclusions are apparent from our study: first, it seems misleading to interpret experimentally obtained NOE distance limits in terms of a single dominant structure. In fact, a variety of stable conformations (clusters of structures) is found. Second, the total number of conformations prevailing in the ensemble is still fairly small. This supports a conclusion previously drawn for \(\beta\)-peptides,[15,26] namely that the accessible portion of the unfolded state consists of significantly fewer conformations than expected from simple combinatorial counts of low-energy conformers.
Experimental Section

With the fully extended structures as a starting point, the carboperoxides 1-5 were simulated in (explicit) chloroform and DMSO (only molecule 1) at 298 K and 1 atm by using the GROMOS96 simulation package\[38\] and the GROMOS biomolecular force field (version 4.5A\[39\]). For simulation and analysis details, see the Supporting Information and ref. \[48\]. The reference (right-handed helical) model structure of the cis tetramer 1 was generated by using MD simulations with all available NOE distance limits implemented as additional restraints. This structure is indeed confirmed to satisfy all distance limits (see Table 1). Model structures of the cis hexamer 2 and cis octamer 3 were constructed in analogy, by assuming the NOE distance limits of the tetramer and augmenting them with the corresponding H–H contacts for the additional residues. This procedure seemed justified, as no experimental data were available to us except for qualitative statements that all cis oligomers share analogous NOE patterns. The left-handed helical model structure for the trans octamer 5 was built by using qualitative information given in ref. \[31\].

Received: February 26, 2004 [Z54114]
Published Online: July 19, 2004

Keywords: computer chemistry · conformation analysis · molecular dynamics · peptide folding · tetrahydrofurans
