Solution Structure of Human Proguanylin

THE ROLE OF A HORMONE PROSEQUENCE*

Thomas Lauber, Philipp Neudecker, Paul Rösch, and Ute C. Marx‡

From the Lehrstuhl für Biopolymere, Universität Bayreuth, Universitätstrasse 30, 95447 Bayreuth, Germany

The endogenous ligand of guanylyl cyclase C, guanylin, is produced as the 94-amino-acid prohormone proguanylin, with the hormone guanylin located at the COOH terminus of the prohormone. The solution structure of proguanylin adopts a new protein fold and consists of a three-helix bundle, a small three-stranded β-sheet of two NH$_2$-terminal strands and one COOH-terminal strand, and an unstructured linker region. The sequence corresponding to guanylin is fixed in its bioactive topology and is involved in interactions with the NH$_2$-terminal β-hairpin: the hormone region (residues 80–94) partly wraps around the first 4 NH$_2$-terminal residues that thereby shield parts of the hormone surface. These interactions provide an explanation for the negligible bioactivity of the prohormone as well as the important role of the NH$_2$-terminal residues in the disulfide-coupled folding of proguanylin. Since the ligand binding region of guanylyl cyclase C is predicted to be located around an exposed β-strand, the intramolecular interactions observed between guanylin and its prosequence may be comparable with the guanylin/receptor interaction.

Activation of intestinal guanylyl cyclase C (GC-C)[1] results in an increased level of intracellular cGMP. Subsequent activation of the cystic fibrosis transmembrane conductance regulator leads to increased secretion of fluid and electrolytes into the intestinal lumen. The first identified intestinal peptide hormone regulating GC-C activity and thus electrolyte and fluid secretion in the intestine is termed guanylin (1, 2).

In circulation, unlike other regulatory peptides, guanylin predominantly occurs as the corresponding 94-amino-acid prohormone, with the hormone located at its very COOH terminus (3). Human proguanylin exhibits only negligible GC-C-activating potency (4), and the *in vivo* active hormone is thought to be

the 15-residue fragment proguanylin (residues 80–94) as originally isolated. This fragment would be released from the prohormone by cleavage of the Asp-79–Pro-80 bond (4). This cleavage, however, also occurs under conditions applied during the original isolation procedure (1, 4), and it is thus still unclear whether Asp-79–Pro-80 is the relevant *in vivo* processing site. For its biological activity, *i.e.* GC-C binding and activation, the hormone requires its two disulfide bonds to be connected in a 1–3/2–4 pattern (5). This unique cysteine connectivity, however, allows formation of two interconvertible topological stereoisomers (6), with only one of them (A-form) showing biological activity (7). It has been shown that the prohormone sequence is essential for the disulfide-coupled folding of the hormone (4) since oxidative folding of reduced fragment proguanylin (residues 80–94) almost completely results in the formation of two inactive disulfide isomers. This crucial role of the hormone prosequence for proper folding was also demonstrated for the closely related uroguanylin prohormone (8, 9).

Unfortunately, only little structural information on prohormones in general is available, and apart from different miniproinsulin mutants, no three-dimensional structure of any prohormone has been deposited in the Protein Data Bank (PDB) yet. Thus, the crucial questions are still open: What is the overall topology of the prohormone? Does the hormone part of proguanylin (residues 80–94) show structural features comparable with those of the active hormone, *i.e.* is its topological isomerism already present in proguanylin? How can the negligible GC-C-activating potency of the prohormone be explained? Why are the NH$_2$-terminal residues important for disulfide-coupled folding? Also, is the proposed processing site to release the hormone exposed to allow attack of a putative proteinase?

In this study, we report the solution structure of proguanylin. In addition, by comparison of the intramolecular interactions found within the proguanylin structure with the proposed ligand binding region of the extracellular domain of the guanylin receptor GC-C, we suggest a mechanism of the guanylin/GC-C interaction.

EXPERIMENTAL PROCEDURES

Sample Preparation—Recombinant proguanylin was expressed as thioredoxin fusion protein in *Escherichia coli* AD494 (DE3) (Novagen, Madison, WI) and purified as described (10). For uniform (>95%) ^{15}N and $^{13}C/^{15}N$ labeling, recombinant proguanylin was isolated and purified from *E. coli* cultures grown in M9 minimal medium enriched with $^{15}NH_4Cl$ or $^{15}NH_4Cl/^{13}C$-glucose using the same purification protocol. As the expression strain lacks the capability to synthesize leucine, ^{15}N-labeled leucine was added to the medium according to the manufacturer's instructions (Novagen). For NMR sample preparation, freeze-dried protein was dissolved to final concentrations of 0.8–1.1 mM in 50 mM potassium phosphate buffer containing 10% D$_2$O (v/v), pH 6.0, or phosphate-buffered D$_2$O (99.994%), pH 6.0.

NMR Spectroscopy—All NMR experiments were performed at either 293 or 298 K on Bruker Avance400 or DRX600 NMR spectrometers (Bruker, Karlsrue, Germany) equipped with inverse $^1H/^{13}C/^{15}N(/^{31}P)$ triple or quadruple resonance probes with pulsed-field gradient capa-

The atomic coordinates and structure factors (code 1o8r) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

‡ To whom correspondence should be addressed. Tel.: 49-921-552048; Fax: 49-921-553544; E-mail: Ute.Marx@uni-bayreuth.de.

[1] The abbreviations used are: GC-C, guanylyl cyclase C; GC-C$_{ECD}$, extracellular ligand binding domain of GC-C; PDB, Protein Data Bank; TOCSY, total correlation spectroscopy; NOE, nuclear Overhauser effect, also used for NOESY cross peak; NOESY, nuclear Overhauser spectroscopy; HSQC, heteronuclear single quantum coherence; HMQC, heteronuclear multiple quantum correlation spectroscopy; r.m.s.d., root mean square deviation.

bilities except for the ^{13}C-NOESY-HSQC, which was recorded on a Bruker Avance800 NMR spectrometer (Bruker, Zürich, Switzerland) equipped with an inverse ^{1}H/^{13}C/^{15}N triple resonance cryoprobe with pulsed-field gradient capabilities. In addition to the homonuclear ^{1}H two-dimensional COSY, TOCSY, and NOESY spectra of 2.4 mM natural proguanylin in 0.5 ml of H$_2$O/D$_2$O (9:1, v/v, pH 4.0) or D$_2$O recorded previously (4), another two-dimensional NOESY spectrum (150-ms mixing time) of natural proguanylin was recorded at 293 K in 0.5 ml of 50 mM potassium phosphate (containing 10% D$_2$O, v/v, pH 6.0), the buffer conditions that were also used for the following heteronuclear two-dimensional and three-dimensional NMR experiments performed on samples of uniformly ^{15}N- or ^{13}C/^{15}N-labeled recombinant proguanylin: ^{15}N-HSQC (11), HNHA (12, 13), ^{15}N-TOCSY-HSQC (14) with DIPSI-2rc (15) (80-ms mixing time), ^{13}C-constant time HSQC (CHSQ) (16), HNCO and HNCA (17, 18), HNCACB (19), CBCA(CO)NH (20), HBHA(CBCACO)NH (21), and H(CCO)NH and C(CO)NH (22) with DIPSI-2 (23) (16-ms mixing time) for the backbone and aliphatic side chain resonance assignment, as well as ^{15}N-NOESY-HSQC (24) (120-ms mixing time), ^{13}C-NOESY-HSQC (25) (150-ms mixing time), and three-dimensional ^{15}N-HMQC-NOESY-HSQC (26, 27) (150-ms mixing time). Side chain ^{1}H resonance assignments of the 8 leucine residues were obtained exclusively from the ^{15}N-edited spectra; a ^{13}C/^{15}N F2-filtered two-dimensional NOESY spectrum (28) (150-ms mixing time) provided additional NOESY data for these residues. Slowly exchanging amide protons were identified from a series of ^{15}N-HSQC spectra every 30 min after dissolving freeze-dried uniformly ^{15}N-labeled proguanylin in D$_2$O. Amide protons still detectable after 150 min were assumed to be involved in hydrogen bonding. D$_{NH}$ residual dipolar couplings were obtained from the observed doublet splitting of uniformly ^{15}N-labeled proguanylin weakly aligned by addition of 21 mg/ml Pf1 filamentous phages (29) in in-phase/antiphase spectra (30) by subtracting the ^{1}J$_{NH}$ scalar coupling contribution as observed in reference spectra of the unaligned sample. {^{1}H}^{15}N NOE values were measured using the pulse sequence of Dayie and Wagner (31) with a relaxation delay of 6 s; the evolution time increments in the indirect dimension of the spectra with and without proton saturation by applying a train of 120° high power pulses for the final 3 s of the relaxation delay were recorded alternately in a single combined experiment. Both D$_{NH}$ residual dipolar couplings and {^{1}H}^{15}N NOE values were averaged over two independent data sets. The H$_2$O resonance was suppressed with excitation sculpting (32), a binomial 3-9-19 WATERGATE sequence (33) with water flip-back (34), and gradient coherence selection (35) in the homonuclear ^{1}H experiments, the amide-detected experiments except for the ^{15}N-TOCSY-HSQC, and the remaining experiments, respectively.

Quadrature detection in the indirectly detected dimensions was obtained by the States-TPPI method (36, 37) or by the echo-antiecho method (38) if coherence selection with gradients was employed (35, 39). Carbon pulses were typically applied as band-selective rectangular, BURP (40), or Gaussian cascade (41) pulses for discriminating aliphatic and carbonyl resonances (42). WALTZ-16 (43), GARP (44), and MLEV-4 expanded off-resonant Gaussian pulse cascades (45) were employed for broadband ^{1}H, broadband ^{13}C or ^{15}N, and band-selective ^{13}C heteronuclear decoupling, respectively. Multidimensional NMR data were extrapolated to up to twice the acquisition time by linear prediction in one of the ^{13}C or ^{15}N indirect dimensions, apodized by multiplication with squared sine bells shifted by 60–90°, and extended to at least twice the length by zero-filling in all dimensions prior to Fourier transformation. The baseline in the direct dimension was corrected using a model-free algorithm (46). The NMR data were processed using in-house written software (47) and analyzed with the program packages NMR-View (48) and NDEE (SpinUp Inc., Dortmund, Germany). ^{1}H chemical shifts were referenced with respect to external 2,2-dimethyl-2-silapentane-5-sulfonic acid in D$_2$O, and ^{13}C and ^{15}N chemical shifts were referenced indirectly (49).

Structure Calculation—Distance restraints for structure calculation were taken from the ^{15}N-NOESY-HSQC spectrum for NOEs involving amide protons and from the ^{13}C-NOESY-HSQC spectrum and homonuclear NOESY spectra for NOEs between aliphatic and aromatic protons. ^{3}J(H^N,H$^\alpha$) coupling constants were calculated from cross-peak to diagonal peak intensity ratios in the HNHA spectrum and corrected by a factor of 1.05 to account for the different relaxation rates of the in-phase and the antiphase component (12), as well as from line-shape analysis of the antiphase cross-signal splitting in a high digital resolution double quantum filtered (DQF)-COSY spectrum using a Lorentzian function for peak fitting.

The non-trivial unambiguous NOESY cross-peaks used for structure calculation were classified manually according to the relative intensi-

ties and converted into upper distance constraints: strong, 2.7 Å; medium, 3.5 Å; and weak, 5.0 Å. ^{3}J(H^N,H$^\alpha$) coupling constants were converted into φ-angles according to the Karplus equation as described (50). All 11 proline residues were considered to have *trans* conformation as evidenced by characteristic Cβ and Cγ chemical shifts (51), unambiguous Hα(i) Pro-Hδ(i+1) NOE signals in the ^{13}C-NOESY-HSQC, and two-dimensional NOESY spectra and the fact that Hα(i) Pro-Hα(i+1) and HN(i) Pro-Hα(i+1) connectivities were absent (52). Disulfide bonds were taken into account by one NOE distance restraint d$_{SS}$ = 2.02 ± 0.10 Å) for each disulfide bond. Hydrogen bonds were selected as described (50, 53) and introduced into the calculation by two distance restraints each (54).

Structures were calculated with X-PLOR 3.851 (50, 55) using a three-stage simulated annealing protocol (56) with floating assignment of prochiral groups (57). To improve the stereochemical quality of the structures, a modified conformational data base potential for backbone and side chain dihedral angles was used (58, 59). The simulated annealing was followed by 1200 steps of energy minimization, with the final 1000 steps without conformational data base potential.

Ambiguities in the NOE cross-peak assignments were solved in an iterative approach using several rounds of structure calculations with subsequent distance analysis. Dihedral angle and hydrogen bond constraints were introduced in later rounds of the structure calculation. Of the 120 structures resulting from the final round of structure calculation, the 40 structures with the lowest overall energies and experimental restraint violations were selected for refinement using additional restraints from residual dipolar couplings: a previously described grid search procedure (53) was applied to obtain the axial component and rhombicity of the alignment tensor using the tensor components of the structures calculated without residual dipolar couplings as starting values. The selected structures were refined using the value of the alignment tensor with the best agreement between measured and calculated residual dipolar couplings. Refinement comprised simulated annealing with a harmonic X-PLOR potential energy term for the dipolar couplings included in the target function of the simulated annealing protocol during the cooling stages. The force constant of the alignment tensor was increased gradually from 0.01 to 1.0 kcal mol^{-1} Hz^{-2}. After refinement, the 30 structures with the lowest overall energies and restraint violations were used for further characterization. The geometry of the structures as well as elements of secondary structure were analyzed using PROCHECK (60) and MOLMOL (61). MOLMOL (61) was used for visualization of the structure data.

Data Bank Accession Numbers—The coordinates and NMR restraints have been deposited in the Protein Data Bank (PDB accession number: 1o8r), and chemical shift values and coupling constants have been deposited in the BioMagResBank (accession number: 5603).

<div align="center">RESULTS</div>

Structure Determination—Based on an almost complete sequence-specific assignment of the backbone (Fig. 1) as well as the side chain resonances of recombinant proguanylin (only the spin system of Leu-47 could not be assigned due to severe spectral overlap), a total of 700 experimental restraints were derived from the NMR spectra. All distance restraints are intramolecular, as natural (*i.e.* proguanylin isolated from human blood ultrafiltrate) as well as recombinant proguanylin are monomeric in solution at concentrations used for the NMR experiments (10).

The experimental restraints consist of 591 distance constraints that were derived from the NOE data in an iterative manner, 30 dihedral angle constraints, 24 hydrogen bonds, and 55 residual dipolar couplings (Table I). 457 NOEs detected in the ^{15}N-NOESY-HSQC could be used as distance restraints, and 65 distance restraints could be obtained from the ^{13}C-NOESY-HSQC spectrum. Additional experimental restraints were derived from the two-dimensional NOESY and COSY spectra of natural proguanylin, providing 69 non-redundant NOEs and 11 ^{3}J(H^N,H$^\alpha$) coupling constants. These restraints were particularly valuable, as they mainly involved long range interactions between aromatic and/or aliphatic side chains that form a hydrophobic core. From the family of 120 structures of the final round of calculations, the 40 structures with the lowest overall energies were chosen for refinement using 55

FIG. 1. **^{15}N HSQC spectrum of ^{15}N-labeled proguanylin.** The almost complete resonance assignment is indicated. Peaks corresponding to side chains of the same Asn or Gln residue are connected by *horizontal lines*. Aliased peaks are colored *gray*. The spectrum was recorded at 293 K and 600 MHz (1.1 mM protein in 50 mM potassium phosphate/10% D_2O, pH 6.0).

TABLE I
Summary of structure calculation

Except for the experimental restraints all values are average values over the 30 final structures in the form average value ($\pm$ standard deviation).

Experimental restraints for final structure calculation	
Interresidual NOEs	547
Sequential ($\lvert i - j \rvert = 1$)	257
Medium range ($\lvert i - j \rvert \leq 5$)	154
Long range ($\lvert i - j \rvert > 5$)	136
Intraresidual NOEs	44
Dihedral angle restraints	
From $^3J(^\alpha H, ^N H)$	30
Hydrogen bonds	24
Restraints from dipolar couplings	55

Molecular dynamics statistics	
Average energy (kcal/mol)	
E_{tot}	183.69 ($\pm$ 5.98)
E_{bond}	5.32 ($\pm$ 0.23)
E_{angles}	141.46 ($\pm$ 2.22)
$E_{improper}$	14.05 ($\pm$ 0.56)
E_{repel}	12.49 ($\pm$ 2.67)
E_{NOE}	4.39 ($\pm$ 1.19)
E_{cdih}	0.14 ($\pm$ 0.097)
E_{dipo}	5.84 ($\pm$ 1.31)
r.m.s.d. from ideal distances (Å)	
NOE	0.012 ($\pm$ 0.0015)
Bonds	0.0048 ($\pm$ 0.00039)
r.m.s.d. from ideal angles (deg)	
Bond angles	1.20 ($\pm$ 0.22)
Improper angles	0.45 ($\pm$ 0.0057)
r.m.s.d. from dipolar couplings (Hz)	
H-N couplings	0.65 ($\pm$ 0.31)

Atomic r.m.s.d. differences from the average structure		
	Backbone	Heavy atoms
	Å	Å
Overall[a]	−1.27	1.89
Residues 1–21, 46–94	0.69	1.14
Regular secondary structure[b]	0.40	0.90
SA_{RDC} *vs.* SA_{noRDC}[c]	0.59	0.68

[a] Calculated for the final set of 30 structures (residues 1–94).

[b] Calculated for the final set of 30 structures (residues 2–4, 9–10, 12–17, 54–62, 66–78, 91–92).

[c] SA_{RDC}, average structure of the final 30 structures calculated with dipolar couplings; SA_{noRDC}, average structure of 30 structures calculated without dipolar couplings (calculated for residues 1–21, 46–94). RDC, residual dipolar couplings.

additional restraints from residual dipolar couplings. Inclusion of these additional restraints increased the precision of the 30 lowest energy structures: the r.m.s.d. of the backbone heavy atoms of elements with regular secondary structure decreased from 0.43 to 0.40 Å. The r.m.s.d. between the average structures of both calculations is 0.59 Å, showing that the orientation of the N-H vectors agree with the overall backbone topology derived from conventional NMR data. From the 40 structures used for refinement, the 30 lowest energy structures were chosen to represent the solution conformation of proguanylin (Fig. 2). These structures show only small deviations in overall energies and r.m.s.d. values from ideal geometry (Table I). This is also confirmed by the PROCHECK analysis (60), showing favorable backbone conformations for almost all residues of the 30 structures with 80.5% of the non-glycine and non-proline residues in the most favored regions and 18.5% in additional allowed regions of the Ramachandran plot. The structures show no distance violations greater than 0.13 Å and no dihedral angle restraint violations greater than 5°.

The best fit backbone superposition of the 30 lowest overall energy structures of proguanylin (Fig. 2) shows a globularly folded protein with well defined elements of regular secondary structure. The r.m.s.d. of the converging structures is 1.27 Å for the backbone heavy atoms and 1.89 Å for all heavy atoms. Omission of the unstructured central linker sequence (residues 22–45) results in lower r.m.s.d. values of 0.69 and 1.14 Å, and elements of regular secondary structure show r.m.s.d. values of 0.40 and 0.90 Å, respectively (Table I).

Description of the Proguanylin Structure—The solution structure of proguanylin consists of a three-helix bundle and a

FIG. 2. **Stereoview of the best-fit backbone superposition of the 30 final proguanylin structures.** Helices (*red*), β-sheet structure (*green*), and regions with low {^{1}H}^{15}N heteronuclear NOE values (*gray*) are indicated. Side chains forming the hydrophobic core (*yellow*) are labeled by residue number (Phe-10, Val-15, Leu-18, Phe-53, Leu-57, Leu-60, Leu-73). The fit is performed for the Cα atoms of the regular secondary structure (residues 2–4, 9–10, 12–17, 54–62, 66–78, 91–92) (figure generated with MOLMOL (61)).

small triple-stranded antiparallel β-sheet of two strands at the NH_2 terminus (Thr-2 to Gln-4, Ser-9 to Phe-10) and one strand at the very COOH terminus (Cys-91 to Thr-92; Figs. 2 and 5A).

FIG. 3. **Atomic r.m.s.d. from the average structure (*A*), distribution of NOEs (*B*), and {¹H}¹⁵N NOE values (*C*).** The r.m.s.d. values for the backbone and all heavy atoms are displayed as *filled* and *open circles*, respectively. The average structure of proguanylin was calculated for elements of regular secondary structure (residues 2–4, 9–10, 12–17, 54–62, 66–78, 91–92). Intraresidual (*black*), sequential (*medium gray*), medium range (*dark gray*), and long range (*light gray*) NOEs are displayed as *vertical stacked bars*. {¹H}¹⁵N NOE values at 600 MHz are displayed as *vertical bars* with *error bars on top*.

FIG. 4. **Hα secondary chemical shifts of proguanylin residues 82–94 and the guanylin A- and B-isomers.** The Hα secondary chemical shifts were calculated using the random coil values given by Wishart *et al.* (73). Chemical shifts of the guanylin A-form (PDB accession number 1gna) and B-form (PDB accession number 1gnb) are taken from Skelton *et al.* (6).

The relative orientation of these structural elements is fixed by a hydrophobic core (Phe-10, Val-15, Leu-18, Phe-53, Leu-57, Leu-60, and Leu-73) whose rigidity is reflected by a high number of long range NOEs (Figs. 2 and 3*B*). Three regular α-helices (Leu-12 to Lys-17, Leu-57 to Lys-62, Ala-66 to Glu-78) are evidenced by several αN(i, i+4) NOEs with helix 2 preceded by a short 3₁₀-helix turn from Pro-54 to Glu-56.

The linker region (Gln-22 to Pro-45) connecting the first two helices is less well defined (Fig. 2), as evidenced by high local r.m.s.d. values (Fig. 3*A*), a relatively low number of detectable NOEs (Fig. 3*B*), and a rapid solvent exchange of the amide protons. The values of the {¹H}¹⁵N heteronuclear NOEs for the amide protons of Gln-22 to Val-44 are all smaller than 0.6 with the lowest value of 0.08 for Ile-38, clearly indicating a significantly increased inherent flexibility of the N-H vectors on a pico- to nanosecond time scale (Fig. 3*C*) (62).

DISCUSSION

The roles of the prosequences for the disulfide-coupled folding and biological activity of the related peptide hormones guanylin and uroguanylin have been investigated previously (4, 8, 9), but the oligomerization state of both prohormones was unclear. In a recent study, we showed by analytical ultracentrifugation that proguanylin is monomeric in solution up to the millimolar range (10), the concentration used for the NMR

experiments. This result evidenced that all distance restraints that were derived from the NMR data are intramolecular. Additionally, we could show that natural and recombinant proguanylin that were both used for structure determination feature identical biophysical and biochemical properties, although recombinant proguanylin possesses an additional Gly and Pro residue at its NH₂ terminus (10).

Relevance of the NH₂-terminal Residues in the Disulfide-coupled Folding of Proguanylin—A major function of the guanylin prosequence is the disulfide-coupled folding of this peptide hormone (4). *In vitro* folding experiments with proguanylin lacking the first 31 amino acids almost exclusively resulted in products with non-native disulfide connectivities (4). A similar observation was made for the related uroguanylin precursor protein, where correct oxidative folding is prevented on truncation of the first 6 residues *in vitro* and *in vivo* (9). Residues Thr-2 to Gln-4 form the central strand of the triple-stranded β-sheet connecting the termini (Figs. 2, 5*A*, and 6), and truncation of the first 6 residues corresponds to deletion of the first β-strand, leading to loss of the interaction between the termini. These tertiary contacts have been shown to be sufficient for the proper conformation of the hormone part (see below), which is likely to be a requisite for generation of the correct disulfide connectivities.

In Its Precursor Protein, Guanylin Adopts the A-form Topology—The originally isolated hormone guanylin is a 15-amino-acid peptide corresponding to proguanylin (residues 80–94). It possesses two disulfide bonds in a 1–3/2–4 connectivity that are essential for its bioactivity (1, 5). Although this disulfide pattern is unique, guanylin is able to form two interconverting topological stereoisomers (6), with only one of them, the A-form, showing significant biological effects (7).

The two isomers are defined by the relative orientation of the loop between the central cystein residues (corresponding residues of the prohormone: Cys-86 and Cys-91) toward the plane formed by the disulfide bonds (6). Both isomers can be distinguished by their distinct sets of spin systems (6, 7). For residues 80–94 of proguanylin, only one set of spin systems could be identified in the NMR spectra with chemical shifts similar to those observed for the A-isomer (Fig. 4): for residues exhibiting the most pronounced differences between the Hα secondary chemical shifts of guanylin isomers A and B, the corresponding residues of proguanylin are more similar to the A-isomer. The same applies for the amide proton shifts (data not shown). Noticeable chemical shift differences between both guanylin isomers and proguanylin residues 82–84 (upfield shift) and 91–93 (downfield shift) are due to secondary structural ele-
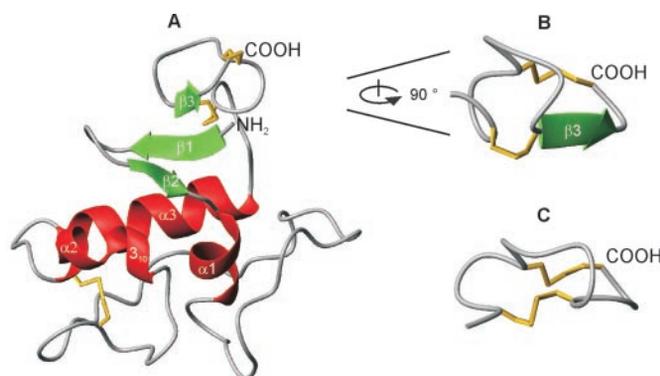
FIG. 5. **Schematic representation of the proguanylin structure.** *A*, schematic drawing of a representative structure of proguanylin indicating elements of regular secondary structure (for elements of regular secondary structure, the same color code is used as in Fig. 2). Disulfide bonds between residues 48–61, 83–91, and 86–94 are displayed in *yellow*. *B*, A-isomer topology of the hormone region of proguanylin (residues 80–94). As compared with the view in *A*, the molecule is rotated by 90° around a vertical axis and zoomed. *C*, guanylin A-isomer (6) (PDB accession number 1gna) (figure generated with MOLMOL (61)).

FIG. 6. **Interaction of the proguanylin hormone region with the NH₂ terminus.** In *A*, a representative structure of proguanylin is shown. Compared with the view in Fig. 5*A*, the molecule is rotated by 45° around a vertical axis. *B*, top view. Helices and the NH₂-terminal β-sheet are colored *red* and *green*, respectively. Residues Val-1 to Gln-4 are displayed as space-filled atoms, and residues 80–94 are shown in stick representation (*blue*; space fill is indicated) (figure generated with RasMol (74)).

FIG. 7. **Sequence alignment (75) of proguanylin and prouroguanylin.** Identical residues are displayed in *red*, conserved and semiconserved substitutions are colored *green* and *blue*, respectively. The experimentally determined (*A*) and predicted (76) (*B*) secondary structures are indicated.

ments, *i.e.* a helical turn (residues 81–84) and the COOH-terminal β-strand (residues 91–92) (Figs. 4 and 5).

Since the NMR spectra show several additional peaks of weaker intensity (Fig. 1), these resonances were examined according to a second hormone topology. However, these additional resonances are due to *cis-trans* isomerization of the 6 proline residues present in the unstructured linker region and to small conformational differences in disulfide bond geometry. Therefore, it is evident that the topological isomerism found for the mature peptide is not present in the precursor protein. In the experimentally determined proguanylin structure, the orientation of the residues corresponding to the central loop region (Ala-87 to Ala-90) is similar to that found for the A-isomer of guanylin (Fig. 5, *B* and *C*): the r.m.s.d. value between the backbone heavy atoms of proguanylin residues 82–94 (Fig. 5*B*) and the guanylin A-form (Fig. 5*C*; PDB accession number 1gna) is 1.62 Å, whereas the r.m.s.d. value for the guanylin B-form (PDB accession number 1gnb) is 5.09 Å.

The A-isomer topology is stabilized in a way that the hormone region of proguanylin partly wraps around the NH₂-terminal β-strand, with residues Val-1 and Val-3 positioned below the plane formed by the disulfide bonds (Cys-83–Cys-91/Cys-86–Cys-94) (Fig. 6). As a result, in addition to the β-sheet typical backbone/backbone interactions between the hormone part (residues 80–94) and the NH₂ terminus, further backbone/backbone, backbone/side chain, and side chain/side chain interactions are present (Fig. 6), as verified by several NOE connectivities, mainly involving residues Val-1 to Gln-4, Cys-86, and Ala-90 to Cys-94. Assuming a hypothetical B-form topology, similar interactions between the hormone part and the prosequence would be sterically unfavored. Thus, the interactions between the termini together with structural features of the prosequence seem to be a commensurate condition for the stabilization of the A-form topology within proguanylin (Fig. 6).

Possible Causes for the Missing Bioactivity of Proguanylin— Although proguanylin residues 80–94 exhibit the bioactive A-isomer topology, the prohormone is almost inactive with respect to GC-C activation (4) due to a significantly decreased receptor binding affinity (63). In contrast, the 32-amino-acid fragment proguanylin (residues 62–94) is able to bind to GC-C with an affinity similar to the biologically active proguanylin (residues 73–94) (63, 64) but is inactive in a T84/cGMP assay (4). As outlined above, the first NH₂-terminal residues of proguanylin including the first β-strand shield part of the hormone surface (Fig. 6), which is therefore not accessible for contacts to GC-C, explaining the significantly decreased receptor binding affinity. This intramolecular interaction is not possible in the NH₂-terminally truncated fragment proguanylin (residues 62–94), which is therefore able to bind to GC-C. Thus, its missing activating potency might be due to sterical properties of the ligand/receptor interaction: for the natriuretic peptide clearance receptor (CNPR) with a ligand binding domain homologous to GC-C, the ligand binds to a cleft formed by the two monomer halves of the dimeric receptor (65). Assuming a similar situation for GC-C, which is, like CNPR, an oligomer prior to ligand binding, the conformational changes necessary for receptor activation might be influenced by the size of the ligand. Further proof of this hypothesis, however, requires structure determination of the guanylin/GC-C complex.

Interaction of Guanylin with Its Receptor GC-C—The extracellular ligand binding domain of GC-C (GC-C$_{ECD}$) is proposed to adopt a periplasmic binding protein (PBP) type I-like fold (66) and is predicted (67) to be structurally similar to the dimerized ligand binding domains of the natriuretic peptide clearance receptor (CNPR; PDB accession number 1jdp) and the guanylyl cyclase-coupled atrial natriuretic peptide receptor (PDB accession number 1dp4). Mapping of the known biochemical data (68–70) on a structural model of GC-C, which is based on these two receptor structures (data not shown), restricts the ligand binding region to a sequence around an exposed and accessible β-strand. Since in its prohormone form guanylin

binds to the NH₂-terminal residues and forms an intramolecular triple-stranded β-sheet (Figs. 5 and 6), it is likely that the same kind of interaction is present in the binding of guanylin to its receptor.

Analysis of Possible Proteinase Cleavage Sites—The native processing site to release the bioactive hormone from its prosequence could not be identified so far as the originally isolated and bioactive 15-residue guanylin can be released from the prohormone by acidic cleavage of an Asp-Pro bond (4), a process that also occurs under conditions applied during the original isolation procedure (1, 4). In the proguanylin structure, this Asp-79–Pro-80 bond is located in the exposed linker region between helix 3 and residues 80–94 (Fig. 5), allowing attack of a potential proteinase without unfolding of the protein. However, no such proteinase has been identified yet. Processing of proguanylin by trypsin digestion results, among others, in the COOH-terminal fragment proguanylin (residues 73–94), which possesses the same biological activity as proguanylin (residues 80–94) (64). The corresponding cleavage site is located at the outside of helix 3 and partly solvent-accessible. This may well mean that Arg-72–Leu-73 is the relevant *in vivo* processing site.

Proguanylin Adopts a New Protein Fold—Only few structural studies on prohormones are published, and of these, only the three-dimensional structures of mini-proinsulin mutants (PDB accession numbers: 1efe, 1sjt, 1sju) have been deposited in the PDB. Analyzing the sequence of proguanylin with 3D-PSSM (67), a program that is able to recognize remote protein sequence homologues as well as structural relationships, no structural homologue with a significant score was found. A DALI protein structure comparison search (71) with the experimentally determined structure of proguanylin came up without a result, strongly supporting the suggestion that proguanylin adopts a new protein fold.

The Related Prouroguanylin Likely Adopts a Fold Similar to Proguanylin—For the related peptide hormones guanylin and uroguanylin, a similar equilibrium between two topological stereoisomers (A- and B-form) can be observed (6, 72), and both peptides adopt similar three-dimensional structures (guanylin, PDB accession numbers: 1gna, 1gnb; uroguanylin, PDB accession numbers: 1uya, 1uyb). Furthermore, an essential contribution of the prosequences of both hormones to the *in vitro* disulfide-coupled folding was demonstrated (4, 8, 9). Additionally, digestion of prouroguanylin with endoproteinase Arg-C and subsequent high performance liquid chromatography analysis showed that the uroguanylin A-form was released exclusively (8). This is similar to the hormone topology observed in the proguanylin structure (Fig. 5*B*). Alignment of the two protein sequences reveals a sequence identity of 35% (Fig. 7). The main difference is a deletion of 7 amino acid residues in the central sequence stretch, located in the flexible linker region of the proguanylin structure. Furthermore, the predicted secondary structure of prouroguanylin closely resembles the experimentally determined one of proguanylin (Fig. 7). Taking all these findings together, proguanylin is a suitable template for homology modeling of prouroguanylin, resulting in similar three-dimensional structures for both proteins (model not shown).

We conclude that the two reported functions of the guanylin prosequence, *i.e.* inactivation of the hormone region and disulfide-coupled folding, are mediated by interactions between NH₂ and COOH terminus: the interstrand contacts of the small triple stranded β-sheet effectively shield the hormone region and are a commensurate condition for an A-form like hormone topology and also for disulfide-coupled folding. Additional reasons for the missing bioactivity of proguanylin and the frag-

ment proguanylin (residues 63–94), like sterical properties, however, cannot totally be excluded.

REFERENCES

1. Currie, M. G., Fok, K. F., Kato, J., Moore, R. J., Hamra, F. K., Duffin, K. L., and Smith, C. E. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 947–951
2. Forte, L. R. (1999) *Regul. Pept.* **81,** 25–39
3. Kuhn, M., Raida, M., Adermann, K., Schulz-Knappe, P., Gerzer, R., Heim, J. M., and Forssmann, W. G. (1993) *FEBS Lett.* **318,** 205–209
4. Schulz, A., Marx, U. C., Hidaka, Y., Shimonishi, Y., Rösch, P., Forssmann, W. G., and Adermann, K. (1999) *Protein Sci.* **8,** 1850–1859
5. Klodt, J., Kuhn, M., Marx, U. C., Martin, S., Rosch, P., Forssmann, W. G., and Adermann, K. (1997) *J. Pept. Res.* **50,** 222–230
6. Skelton, N. J., Garcia, K. C., Goeddel, D. V., Quan, C., and Burnier, J. P. (1994) *Biochemistry* **33,** 13581–13592
7. Schulz, A., Escher, S., Marx, U. C., Meyer, M., Rösch, P., Forssmann, W. G., and Adermann, K. (1998) *J. Pept. Res.* **52,** 518–525
8. Hidaka, Y., Ohno, M., Hemmasi, B., Hill, O., Forssmann, W. G., and Shimonishi, Y. (1998) *Biochemistry* **37,** 8498–8507
9. Hidaka, Y., Shimono, C., Ohno, M., Okumura, N., Adermann, K., Forssmann, W. G., and Shimonishi, Y. (2000) *J. Biol. Chem.* **275,** 25155–25162
10. Lauber, T., Nourse, A., Schulz, A., and Marx, U. C. (2002) *Biochemistry* **41,** 14602–14612
11. Mori, S., Abeygunawardana, C., Johnson, M. O., and Vanzijl, P. C. M. (1995) *J. Magn. Reson. Series B* **108,** 94–98
12. Vuister, G. W., and Bax, A. (1993) *J. Am. Chem. Soc.* **115,** 7772–7777
13. Zhang, W., Smithgall, T., and Gmeiner, W. H. (1997) *J. Biomol. NMR* **10,** 263–272
14. Zhang, O., Kay, L. E., Olivier, J. P., and Forman-Kay, J. D. (1990) *J. Biomol. NMR* **4,** 845–858
15. Cavanagh, J., and Rance, M. (1992) *J. Magn. Reson.* **96,** 670–678
16. Vuister, G. W., and Bax, A. (1992) *J. Magn. Reson.* **98,** 428–435
17. Ikura, M., Kay, L. E., and Bax, A. (1990) *Biochemistry* **29,** 4659–4667
18. Grzesiek, S., and Bax, A. (1992) *J. Magn. Reson.* **96,** 432–440
19. Wittekind, M., and Mueller, L. (1993) *J. Magn. Reson., Series B* **101,** 201–205
20. Grzesiek, S., and Bax, A. (1992) *J. Am. Chem. Soc.* **114,** 6291–6293
21. Grzesiek, S., and Bax, A. (1993) *J. Biomol. NMR* **3,** 185–204
22. Grzesiek, S., Anglister, J., and Bax, A. (1993) *J. Magn. Reson., Series B* **101,** 114–119
23. Shaka, A. J., Lee, C. J., and Pines, A. (1988) *J. Magn. Reson.* **77,** 274–293
24. Talluri, S., and Wagner, G. (1996) *J. Magn. Reson., Series B* **112,** 200–205
25. Ikura, M., Kay, L. E., Tschudin, R., and Bax, A. (1990) *J. Magn. Reson.* **86,** 204–209
26. Frenkiel, T., Bauer, C., Carr, M. D., Birdsall, B., and Feeney, J. (1990) *J. Magn. Reson.* **90,** 420–425
27. Ikura, M., Bax, A., Clore, G. M., and Gronenborn, A. M. (1990) *J. Am. Chem. Soc.* **112,** 9020–9022
28. Zwahlen, C., Legault, P., Vincent, S. J. F., Greenblatt, J., Konrat, R., and Kay, L. E. (1997) *J. Am. Chem. Soc.* **119,** 6711–6721
29. Hansen, M. R., Mueller, L., and Pardi, A. (1998) *Nat. Struct. Biol.* **5,** 1065–1074
30. Ottiger, M., Delaglio, F., and Bax, A. (1998) *J. Magn. Reson.* **131,** 373–378
31. Dayie, K. T., and Wagner, G. (1994) *J. Magn. Reson., Series A* **111,** 121–126
32. Hwang, T. L., and Shaka, A. J. (1995) *J. Magn. Reson., Series A* **112,** 275–279
33. Sklenar, V., Piotto, M., Leppik, R., and Saudek, V. (1993) *J. Magn. Reson., Series A* **102,** 241–245
34. Grzesiek, S., and Bax, A. (1993) *J. Am. Chem. Soc.* **115,** 12593–12594
35. Schleucher, J., Schwendinger, M. G., Sattler, M., Schmidt, P., Schedletzky, O., Glaser, S. J., Sørensen, O. W., and Griesinger, C. (1994) *J. Biomol. NMR* **4,** 301–306
36. States, D. J., Haberkorn, R. A., and Ruben, D. J. (1982) *J. Magn. Reson.* **48,** 286–292
37. Marion, D., Ikura, M., Tschudin, R., and Bax, A. (1989) *J. Magn. Reson.* **85,** 393–399
38. Kay, L. E., Keifer, P., and Saarinen, T. (1992) *J. Am. Chem. Soc.* **114,** 10663–10665
39. Sattler, M., Schwendinger, M. G., Schleucher, J., and Griesinger, C. (1995) *J. Biomol. NMR* **5,** 11–22
40. Geen, H., and Freeman, R. (1991) *J. Magn. Reson.* **93,** 93–141
41. Emsley, L., and Bodenhausen, G. (1990) *Chem. Phys. Lett.* **165,** 469–476
42. Sattler, M., Schleucher, J., and Griesinger, C. (1999) *Prog. NMR Spectrosc.* **34,** 93–158
43. Shaka, A. J., Keeler, J., Frenkiel, T., and Freeman, R. (1983) *J. Magn. Reson.* **52,** 335–338
44. Shaka, A., Barker, P. B., and Freeman, R. (1985) *J. Magn. Reson.* **64,** 547–552
45. Eggenberger, U., Schmidt, P., Sattler, M., Glaser, S. J., and Griesinger, C. (1992) *J. Magn. Reson.* **100,** 604–610
46. Friedrich, M. S. (1995) *J. Biomol. NMR* **5,** 147–153
47. Schweimer, K. (2000) *Mehrdimensionale NMR Spektoskopie zur Bestimmung der Strukturen des Birkenpollenallergens Bet v 1, des Guillardia theta Rubredoxins und des [2Fe-2S] Ferredoxins aus Halobakterium salinarium.* Doctoral Dissertation, University of Bayreuth, Bayreuth, Germany
48. Johnson, B. A., and Blevins, R. A. (1994) *J. Biomol. NMR* **4,** 603–614
49. Markley, J. L., Bax, A., Arata, Y., Hilbers, C. W., Kaptein, R., Sykes, B. D., Wright, P. E., and Wüthrich, K. (1998) *Pure Appl. Chem.* **70,** 117–142
50. Neudecker, P., Schweimer, K., Nerkamp, J., Scheurer, S., Vieths, S., Sticht, H., and Rösch, P. (2001) *J. Biol. Chem.* **276,** 22756–22763
51. Hill, J. M., Alewood, P. F., and Craik, D. J. (1996) *Biochemistry* **35,** 8824–8835
52. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, pp. 122–125, John Wiley & Sons, Inc., New York
53. Schweimer, K., Hoffmann, S., Bauer, F., Friedrich, U., Kardinal, C., Feller, S. M., Biesinger, B., and Sticht, H. (2002) *Biochemistry* **41,** 5120–5130

54. Kraulis, J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J., and Gronenborn, A. M. (1989) *Biochemistry* **28,** 7241–7257
55. Brünger, A. T. (1993) *X-PLOR*, Version 3.1, a system for X-ray crystallography and NMR, Yale University Press, New Haven, CT
56. Nilges, M. (1993) *Proteins* **17,** 297–309
57. Folmer, R. H., Hilbers, C. W., Konings, R. N., and Nilges, M. (1997) *J. Biomol. NMR* **9,** 245–258
58. Kuszewski, J., and Clore, G. M. (2000) *J. Magn. Reson.* **146,** 249–254
59. Neudecker, P., Sticht, H., and Rösch, P. (2001) *J. Biomol. NMR* **21,** 373–375
60. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Cryst* **26,** 283–291
61. Koradi, R., Billeter, M., and Wuthrich, K. (1996) *J. Mol. Graph.* **14,** 29–32
62. Kay, L. E., Torchia, D. A., and Bax, A. (1989) *Biochemistry* **28,** 8972–8979
63. Garcia, K. C., de Sauvage, F. J., Struble, M., Henzel, W., Reilly, D., and Goeddel, D. V. (1993) *J. Biol. Chem.* **268,** 22397–22401
64. de Sauvage, F. J., Keshav, S., Kuang, W. J., Gillett, N., Henzel, W., and Goeddel, D. V. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 9089–9093
65. He, X., Chow, D., Martick, M. M., and Garcia, K. C. (2001) *Science* **293,** 1657–1662
66. van den Akker, F. (2001) *J. Mol. Biol.* **311,** 923–937
67. Kelley, L. A., MacCallum, R. M., and Sternberg, M. J. (2000) *J. Mol. Biol.* **299,** 501–522
68. Hasegawa, M., Hidaka, Y., Matsumoto, Y., Sanni, T., and Shimonishi, Y. (1999) *J. Biol. Chem.* **274,** 31713–31718
69. Wada, A., Hirayama, T., Kitaura, H., Fujisawa, J., Hasegawa, M., Hidaka, Y., and Shimonishi, Y. (1996) *Infect. Immun.* **64,** 5144–5150
70. Hidaka, Y., Matsumoto, Y., and Shimonishi, Y. (2002) *FEBS Lett.* **526,** 58–62
71. Holm, L., and Sander, C. (1993) *J. Mol. Biol.* **233,** 123–138
72. Marx, U. C., Klodt, J., Meyer, M., Gerlach, H., Rösch, P., Forssmann, W. G., and Adermann, K. (1998) *J. Pept. Res.* **52,** 229–240
73. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) *J. Biomol. NMR* **5,** 67–81
74. Sayle, R. (1995) *RasMol*, V2.6, molecular visualization program, Glaxo Wellcome Research and Development, Stevevage, Hertfordshire, UK
75. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673–4680
76. Jones, D. T. (1999) *J. Mol. Biol.* **292,** 195–202