Research Article

Journal of PeptideScience

Received: 10 November 2008 Revised: 12 December 2008 Accepted: 14 December 2008 Published online in Wiley Interscience: 16 February 2009

(www.interscience.com) DOI 10.1002/psc.1122

Design of stable α-helices using global sequence optimization

Michael Petukhov,[a*] Yoshiro Tatsu,[b] Kazuyo Tamaki,[b] Sachiko Murase,[b‡] Hiroko Uekawa,[b‡] Susumu Yoshikawa,[b‡] Luis Serrano[c] and Noboru Yumoto[b]

The rational design of peptide and protein helices is not only of practical importance for protein engineering but also is a useful approach in attempts to improve our understanding of protein folding. Recent modifications of theoretical models of helix-coil transitions allow accurate predictions of the helix stability of monomeric peptides in water and provide new possibilities for protein design. We report here a new method for the design of α-helices in peptides and proteins using AGADIR, the statistical mechanical theory for helix-coil transitions in monomeric peptides and the tunneling algorithm of global optimization of multidimensional functions for optimization of amino acid sequences. CD measurements of helical content of peptides with optimized sequences indicate that the helical potential of protein amino acids is high enough to allow formation of stable α-helices in peptides as short as of 10 residues in length. The results show the maximum achievable helix content (HC) of short peptides with fully optimized sequences at 5 °C is expected to be ∼70–75%. Under certain conditions the method can be a powerful practical tool for protein engineering. Unlike traditional approaches that are often used to increase protein stability by adding a few favorable interactions to the protein structure, this method deals with all possible sequences of protein helices and selects the best one from them.

Keywords: α-helix; design; sequence optimization; AGADIR; tunneling algorithm

Introduction

Several factors have been shown to be important for the stabilization of α-helices in monomeric peptides, as well as in proteins. These factors include interactions between amino acid side chains [1–3], interactions between charged or polar side chains and the helix macrodipole [4–6], and terminal capping [7]. All these factors have been applied separately in attempts to increase the conformational stability of α-helices in peptides and in natural proteins [8,9]. However, the design of peptide sequences with the optimal exploitation of all these factors cannot be performed by current techniques, even in the case of short peptides, because of the complicated nature of the α-helix wherein an accurate balance of interactions of many kinds controls the helical stability. For example a positively charged amino acid at position i requires that the $i + 3$, $i + 4$ and also the $i - 3$, $i - 4$ positions of the helix are occupied by negatively charged residues that may be unfavorable for formation of a helix if they occur close to the carboxy-terminus because of interactions with the helix macrodipole. The difficulty of the problem grows exponentially with peptide length because of the necessity of accounting for the many factors that affect helical stability and for all possible arrangements of helical segments in the molecule under investigation [10–13].

The number of possible sequences of a peptide with N amino acid residues is as high as 20^N. Thus, it is computationally impossible to calculate the helical content of each possible sequence even for short peptides. To overcome this problem we used the tunneling algorithm for global optimization of multidimensional functions [14]. The main advantage of this approach is that it does not require an examination of all possible sequences to find the best one. The method is simple and robust, requiring calculations only of first derivatives of the goal function. It has been reported that the method was able to identify global minima of many problems with many thousands of local minima.

Unfortunately, all available global optimization techniques are a sort of random walkers covering more or less a significant area of phase space of a problem under investigation. None of them can prove true globality of a found solution. Due to this and also due to imperfectness of theoretical approximations that can be used to predict helix stability it is unlikely any peptide sequences can be rigorously proved to be globally optimized to date and foreseeable future. However, despite inability of theoretical models for guarantee convergence to a true globally optimized peptide sequence, an efficient tool for protein helix optimization is still needed for many protein engineering applications where a bit less optimal sequences can be an adequate replacement of truly globally optimized ones. Creating and testing such a tool on short peptide helices was the main goal of the work. The other goals also include answering to the following basic questions:

* *Correspondence to: Michael Petukhov, Department of Molecular and Radiation Biophysics, Petersburg Institute of Nuclear Physics, RAS, Orlova rosha, Gatchina, 188300, Russia. E-mail: michael.petukhov@gmail.com*

a *Petersburg Institute of Nuclear Physics, the Russian Academy of Sciences, 188300, Gatchina, Russia*

b *National Institute of Advanced Industrial Science and Technology (AIST), Ikeda, Osaka 563-8577, Japan*

c *Icrea Professor, CRG-EMBL Systems Biology Unit, Centre de Regulacio Genomica, Dr Aiguader 88 08003 Barcelona, Spain*

‡ *Part of this work done by Sachiko Murase, Hiroko Uekawa, Susumu Yoshikawa was carried out at Osaka National Research Institute which was integrated to AIST from 2001.*

(i) What is the maximum level of helicity that can be achieved in short peptides containing 20 protein amino acids? (ii) What kinds of sequence motif exhibit close to the optimal arrangement of intrahelical interactions and how does this arrangement depend on the length of the peptide? (iii) If such motifs do exist, how are they related to the helices of natural proteins?

In this study we have attempted to shed some light on these questions using statistical mechanical calculations based on the AGADIR model [15] and several of its more recent modifications [10,12,13,16–19] and the global optimization algorithm [14] for the design of peptide helices. In this study, we have synthesized some theoretically designed peptides with optimized sequences and measured their helical contents by CD.

Materials and Methods

Calculations Based on Statistical Mechanics

The calculations of helical content (the target function for our global optimization procedure) were based on one sequence approximation AGADIR (AGADIR1s), a statistical mechanical approach that accurately predicted the helical properties of several hundred short peptides in water [10,11]. Since, unlike globular protein, short peptides do not have a single stable conformation, AGADIR model accounts for free energy contribution from all possible helical segments in the peptide under consideration as follows. The difference in free energy between the random-coil and helical states for a given segment is calculated as the following summation [12]:

$$\Delta G_{\text{helical}-\text{segment}} = \Delta G_{\text{int}} + \Delta G_{\text{hbond}} + \Delta G_{\text{sc}-\text{sc}} + \Delta G_{\text{electrostat}} + \Delta G_{\text{nonH}} + \Delta G_{\text{macrodipole}}$$

where ΔG_{int} is the summation of the intrinsic propensities of all residues in a given helical segment (including its observed positional dependencies [13,18,19]), that is the loss of conformational entropy required to adopt α-helical dihedral angles; ΔG_{hbond} is the sum of the main-chain–main-chain enthalpic contributions, which include the formation of i, i + 4 hydrogen bonds; $\Delta G_{\text{sc}-\text{sc}}$ sums the net contributions, with respect to the random-coil state, of all side-chain–side-chain interactions located at positions i, i + 3 and i, i + 4 in the helical region; $\Delta G_{\text{electrostat}}$ includes all electrostatic interactions between two charged residues inside and outside the helical segment; ΔG_{nonH} represents the sum of all contributions to helix stability of a given segment from residues that are not in a helical conformation (N and C-capping, Capping Box, hydrophobic staple motif, Schellman motif etc.); $\Delta G_{\text{macrodipole}}$ represents the interaction of charged groups with the helix macrodipole. In AGADIR model the helix content (HC) of a peptide under consideration is calculated as:

$$HC = \frac{\sum e^{-\frac{\Delta G_{\text{helical}-\text{segment}}}{RT}}}{1 + \sum e^{-\frac{\Delta G_{\text{helical}-\text{segment}}}{RT}}};$$

where the sum is over all possible α-helical segments.

In addition to the original AGADIR set of energy parameters [10] we incorporated several modifications of the parameter set of the theory published later [12]. Also the dependence of the intrinsic propensities of amino acids on their positions within helical segments was incorporated, as has been described [13,18,19]; and the energy parameters for those helical segments where formation of a capping box was possible were calculated as described [17]. The dependence of the energy parameters on temperature and pH was included according to Munoz and Serrano [11].

Global Optimization

The Tunneling Algorithm [14] was used in the sequence optimization calculations. It incorporates two main phases: a local minimization phase and a tunneling phase. During the minimization phase, the target function $f(x)$ is minimized by the conjugate gradient method as implemented in Fletcher–Reeves method [20]. During the tunneling phase, the algorithm starts from the vicinity of the point, x^*, reached during the previous phase and search for a zero value of the auxiliary function:

$$T(x, x^*, f^*) = \frac{f(x) - f^*}{(x - x^*)^{2a}(x - x^m)^{2b}}$$

by the modified Newton method [21]. In the above equation x^* and a are the position and the strength of the immobile pole respectively; x^m and b are those of the mobile pole; and $f^* = f(x^*)$. The function $T(x, x^*, f^*)$ has continuous first derivatives and the same set of zeros as the function $f(x) - f^*$ does. The immobile pole is intended to prohibit the algorithm from returning to the point x^*. If the step of the Newton method becomes too small, the algorithm switches on the mobile pole. This pole is intended to eject the optimization process from such points. The aim of the tunneling phase is the identification of a point $x \neq x^*$ where $f(x) \leq f(x^*)$. It is obvious that such a point would be very attractive as a starting point for the subsequent local minimization phase.

Nonconvergence of the tunneling phase during 100 iterations of the algorithm was defined to be the stop condition for termination of the optimization process. Optimizations were performed for four series of peptides of 13, 14, 15, and 16 residues in length. Two modes of optimization were used for each series of peptides: (i) optimization starting from all Gly residues until the stopping condition was met and (ii) a long-run optimization (1 000 000 estimations in all of the helical content), starting each time from random sequences. In all the cases, the sequence of three carboxy-terminal residues was fixed at Gly-Gly-Tyr (see below).

Interpolation of Energy Parameters

If we are to use Tunneling Algorithm for peptide sequence optimization, it is necessary to treat the amino acids of the primary structure as real variables. Therefore, we interpolated all the discrete energy parameters used in the statistical mechanical calculations of the goal function as follows: (i) integers from 1 to 20 were assigned to each type of amino acid; (ii) the energy parameters of the AGADIR system were assigned to these integers on the real axis; (iii) energy barriers of 2.5 kcal/mol were introduced at the mid points between the integers assigned to the amino acids; and (iv) the regular grids of the energy parameters and the barriers were used for one-dimensional and two-dimensional cubic spline interpolations [22]. The splines obtained by this procedure are continuously differentiable functions with well-separated energy minima at the integer points of the real axis where they have both the true values of the AGADIR set of energy parameters and zero gradients.

To avoid the uncertainties that are associated with the tendency of the Tunneling Algorithm to escape from the permitted range

Journal of PeptideScience

of the real axis (from 1 to 20), the following periodical boundary conditions were employed for all points of the real axis:

$$P_{\text{int}}(t_{aa} + n \times 20) = P_{\text{int}}(t_{\text{aa}})$$

where P_{int} is the interpolation value of a parameter, t_{aa} is a variable type of amino acid, and n is an integer.

The executables of the sequence optimization software are available upon a request from the corresponding author.

Synthesis of Peptides and CD Measurements

Peptides with optimized sequences were synthesized on an automated solid-phase peptide synthesizer (PSSM-8, Shimadzu, Kyoto, Japan), using Tenta Gel TG-RAM resin and Fmoc chemistry, with benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate and N-hydroxybenzo-triazole as coupling reagents. Peptides were cleaved from the resin with trifluoroacetic acid and purified by reverse-phase HPLC on a C_{18} column. Fmoc-L-amino acids, reagents for peptide synthesis, and Tenta Gel TG-RAM resin were purchased from Shimadzu (Kyoto, Japan). Fmoc amino acids used were Fmoc-Ala-OH, Fmoc-Arg(Pmc)-OH, Fmoc-Asp(OtBu)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Glu(OtBu)-OH, Fmoc-Gly-OH, Fmoc-Leu-OH, Fmoc-Lys(Boc)-OH, Fmoc-Ser(tBu)-OH, Fmoc-Thr(tBu)-OH, Fmoc-Ser(tBu)-OH, and Fmoc-Tyr(tBu)-OH.

The purity of each peptide was assessed by analytical reverse-phase HPLC on a C_{18} column. All the peptides were obtained as a single major product by peptide synthesis and purification. Stock solutions of the peptides were prepared in water at 2 mM concentrations. Molecular masses were confirmed by mass spectrometry on a time-of-flight mass spectrometer (Kompact MALDI II, Shimadzu/Kratos) with matrix-assisted laser desorption ionization and electrospray ionization mass spectrometry (ESI-MS) (Thermo electron, LC deca XP plus).

CD measurements were made with JASCO spectropolarimeters J-500A and J-820. CD spectra were recorded at 5 °C in a thermostatically controlled cell with a 0.1–cm path length. Each peptide was dissolved in 5 mM sodium phosphate buffer (pH 7.5) at a concentration of approximately 10 μM. Peptide concentrations were calculated by the absorption of tyrosine residue at 280 nm. Then α-helical contents were calculated from mean residue ellipticities measured at 222 nm (θ_{222}) as described by Chen *et al*. [23].

Results and Discussion

We used methods described in the section Materials and Methods to obtain the best helical sequences for four peptide series of 13, 14, 15, and 16 residues length. The three carboxy-terminal amino acids were fixed during the optimization at Gly-Gly-Tyr for the following reasons: (i) the convenience of determination of concentrations from UV-absorbance of Tyr; and (ii) the absence of the aromatic contribution of carboxy-terminal Tyr to CD at 222 nm if it is separated from the rest of a peptide by two flexible Gly residues [24]. In addition, the expected lengths of helical segments of these four peptides (from 10 to 13 residues) correspond approximately to average lengths of protein α-helices that have been observed by protein crystallography [25]. The starting sequences of the variable part of the peptides in the four runs were set to Gly residues for simplicity.

Figure 1 shows the progress of typical short optimizations for the 13-, 14-, 15-, and 16-residue peptides. One can see from the figure that, in the majority of the cases, the method of sequence optimization converged to the best sequence within a few tens of thousand steps, an indication of high enough efficiency of the method. The best sequences obtained from these four short optimizations had a few common features: (i) the sequences of the four amino-terminal residues always corresponded to what is called 'Capping Box' motif, wherein side chains of the first (Thr) and the fourth (Glu) residues can form a specific pattern of hydrogen bonding, with the amide protons of the main chain stabilizing the α-helix [17,26]; (ii) the carboxy-terminal positions were often occupied by positively charged amino acids that can stabilize an α-helix by charge–helix macrodipole interactions; c) salt bridges between side chains of Glu^- and Arg^+ or of Glu^- and Lys^+ at central positions provided another stabilizing factor that was very common in these sequences. Obviously, the number of the salt bridges and their arrangement depended on the length of the peptide affecting availability of central positions for amino acids with opposite charge to that of favorable amino acids located near amino- and carboxy-termini. However, at this point it is still remaining unclear whether or not other types of side chain interactions can compete with salt bridge formation in terms of α-helix stabilization.

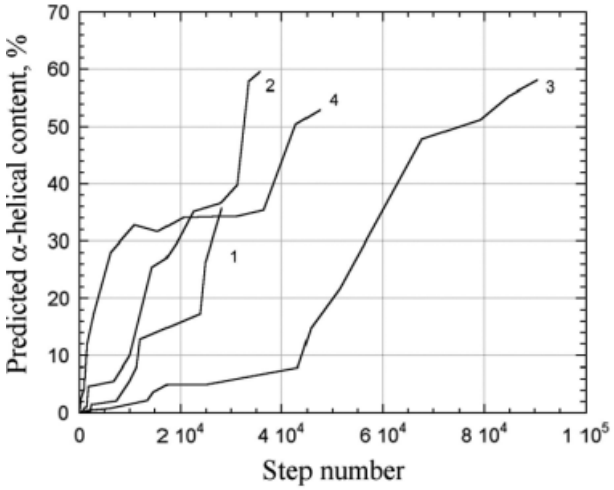

Figure 1. The progress of a typical sequence optimizations for (1) 13-residue peptides; (2) 14-residue peptides; (3) 15-residue peptides, and (4) 16-residue peptides with acetylated amino- and amidated carboxy-termini.

The optimized sequences obtained from these calculations are presented in Table 1 (series A to H). Peptides of this series are predicted to have helical contents in the broad range of 20–60% at 5 °C and pH 7.5. The low temperature was chosen in this study because this is the standard temperature for measurements of helix stability in short monomeric peptides adopted in many papers published in the field. Also the CD measurements at higher helix content provide a better accuracy of the measurements. However, the temperature is quite an important factor for helix sequence optimization since different types of free energy contributions are not equally sensitive for temperature changes. For instance sequence optimization of a protein would certainly require use of proper target temperature for its functioning. It is noteworthy that AGADIR model includes temperature and pH dependencies for all used terms of free energy contributing to helix stability and therefore it can be used without any modifications for sequence optimizations at any target temperature.

Table 1. Sequences and results of MS and CD measurements for series of peptides with optimized sequences[a]

Peptide (sequence)	Exact mass (calc.) (D)	Observed mass $[M+H^+]/[M+2H^+](m/z)$	$-\theta_{222}$ (deg cm^2/dmol)	Helix content (exp.) %
A (TEAEERAKAAGGY-amide)	1350.7	1352	8974	28.9
B (TEAEERAKASGGY-amide)	1366.6	1367	7773	25.0
C (TEQEERERLRKGGY-amide)	1748.9	1751	15 095	47.6
D (TEQEERERARKGGY-amide)	1706.8	1708	14 466	45.6
E (TAREEAERARRKGGY-amide)	1747.9	1749	13 716	42.5
F (TAAEEAERARRKGGY-amide)	1662.9	1663	13 982	43.4
G (TEAEAAELKARLKGGY-amide)	1704.9	1708	12 339	37.7
H (TNEEAARAALAKKGGY-amide)	1647.9	1649	12 995	39.7
O13 (ac-REEEERRRRLGGY-amide)	1745.9	1748	18 430	58.2
O14 (ac-DRKREEEERRRGGY-amide)	1875.9	1878	17 730	55.0
O15 (ac-DRLLEELLRRLLGGY-amide)	1856.1	1858	10 874	33.2
O16 (ac-DEERERLELLLRLGGY-amide)	2001.1	1002[b]	18 298	55.2

[a] The sequences of the peptides were obtained by the sequence optimization algorithm starting from poly-Gly sequences. Helix content was estimated using the standard formula $-\theta_{222}/39\,500(1-2.57/N_{aa})$, where N_{aa} is the number of amino acid residues [23]. The peptides A–H were measured by MALDI-TOF MS and the peptides O13–O16 were measured by ESI-MS.
[b] $[M+2H^+]$

To confirm the validity of our theoretical predictions, we synthesized eight peptides (one pair of best sequences from each of the four optimizations) and measured their helical content by CD. Selection of these peptides sequences served two purposes: (i) it displays real level of single run efficiency of the algorithm to search for a highly optimized peptide sequence; and (ii) it shows that despite all the theoretical approximations the AGADIR predictions for highly optimized sequences are in reasonably good agreement with experiment.

Figure 2 shows the CD spectra of this peptide series under the standard conditions (temperature 5 °C and pH 7.5). Their mean residue ellipticities at 222 nm and calculated helix contents are listed in Table 1. As expected, the CD spectra of all the peptides show typical α-helical shape with helix content ranging from 25 to 47%.

To examine whether the best sequences obtained from these four optimizations were globally optimized, we performed multiple optimization runs for peptides of the same length, starting each time from random sequences (totally 1 000 000 estimations of the goal function for each peptide). The 20 best sequences obtained in these long run optimizations, together with their predicted helicities are presented in Table 2. One can see from this Table that the sequences obtained from the first optimization series (see Table 1) are not listed among the 20 best sequences found in the subsequent optimizations. These results indicate that the Tunneling Algorithm, as well as any other currently available methods for global optimization of multidimensional functions, cannot guarantee convergence to globally optimized sequences. Apart from this reservation, the method showed a high efficiency in the search for very stable α-helical sequences, although maybe not a one corresponding to global minimum of the goal function. Therefore, the method seems clearly adequate for many practical problems of protein and peptide engineering.

It is of interest that the helical stability of the series of peptides presented in Table 2 conforms to a reasonable rank order of best sequences according to its length. The increase in length of a peptide by one residue was associated with an approximately ~3–5% increase in the helical content in the optimized sequences.

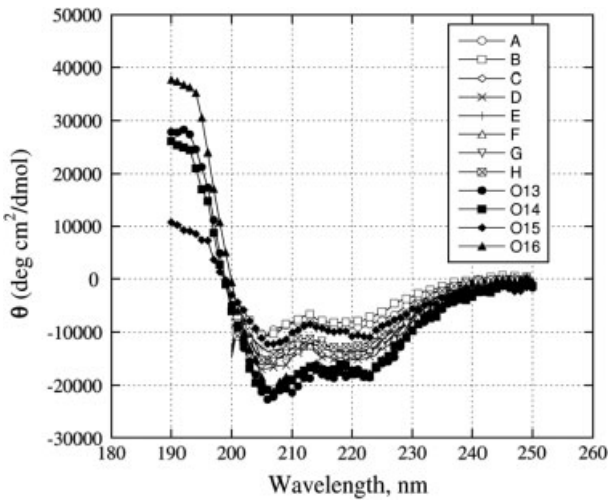

Figure 2. Far-UV CD spectra of synthetic peptides with optimized sequences recorded under standard conditions at 5 °C in 5 mM (TES-NaOH) buffer (pH 7.5) for 13-, 14-, 15-, and 16-residue peptides. The sequences of the peptides are listed in Table 1. The peptide concentrations were approximately 10 μM. The amino-termini were free (peptides A–H) or acetylated (peptides O13–O15) and the carboxy-termini of all peptide were amidated.

These observations indicate that the sequences in Table 2 are not random and probably do represent close to the upper limit of helical stability of peptides of these lengths.

The sequences in Table 2 also have some common features. The most stable peptide helices consist of only a few amino acid types (Leu, Met, Trp, Tyr, Glu, and Arg) having both the good internal helical propensities and high potential for other stabilizing interactions such as side chain–side chains interactions and N- and C-capping interactions. It is of interest that top positions of the peptide series are occupied by poly-Leu and poly-Trp motifs indicating that multitude of the favorable hydrophobic side chain–side chain interactions fully compensates loss of other

Journal of PeptideScience

Table 2. The 20 best sequences obtained by sequence optimization for 13-, 14-, 15-, and 16-residue peptides[a]

Optimized sequences of 13-residue peptides	Fraction of predicted α-helix %	Optimized sequences of 14-residue peptides	Fraction of predicted α-helix%	Optimized sequences of 15-residue peptides	Fraction of predicted α-helix %	Optimized sequences of 16-residue peptides	Fraction of predicted α-helix %
LLLLLLLLLLGGY	79.9	WWWWYYYYMMMGGY	82.0	EELLRRLLLLLLGGY	84.2	LELLLRLLLLLLLGGY	84.3
WWWWWWYMWYGGY	78.0	LLLLLLLLLLLLGGY	81.8	LLLLLLLLLLLLGGY	83.1	LLLLLLLLLLLLLGGY	84.1
TLLLLLLLLLGGY	76.0	EELLRLLLLLLGGY	81.1	LYLLMWLLMMLLGGY	83.0	LEYLLMWLLMMLLGGY	84.0
WWLLYMLIWLGGY	75.3	MYLLMWLLMWLGGY	81.1	LELLLRLLLLLLGGY	82.9	LEELLRRLLLLLLGGY	83.9
DLLLLLLLLLGGY	75.2	WWWWWWWWWWGGY	81.1	WFFLMMWLLMWLGGY	80.6	QWWWYYYYMMMWLGGY	83.1
SLLLLLLLLLGGY	74.2	SLLLLLLLLLLGGY	77.3	WWLWYYMYMWYWGGY	80.5	MLEELLRLLLLLLGGY	82.1
REEEERRRRLGGY	73.6	RLRELLERLLRGGY	74.8	TLLLLLLLLLLLGGY	80.4	FRWYLEYWLRMWLGGY	82.0
NLLLLLLLLLGGY	73.4	SREEEERRRRLGGY	74.5	LLEELLRLLLWLGGY	79.3	MIEELLRLLLLLLGGY	81.8
SREEEERRRRGGY	69.5	SLVELLLLLLLGGY	74.3	SLLELLLLLLLLGGY	77.6	FMAYLWYWYYMYMGGY	81.7
SLLEELLRRLGGY	67.9	**DRKREEEERRRGGY**	74.3	DALLELLLRLLLGGY	77.3	SLLLLLLLLLLLLGGY	80.5
DLLLELLLRLGGY	67.7	DKKREEEERRRGGY	74.1	SLLLELLLRLLLGGY	77.3	SWWWYMYYWYMMYGGY	79.9
DLLEELLRRLGGY	67.4	DRLLEELLRRLGGY	74.0	**DRLLEELLRRLLGGY**	77.2	TMLLELLLLLLLLGGY	79.6
DAEEELRRRLGGY	67.3	DRQRLEEELRRGGY	73.9	DELLERLLRLLLGGY	77.1	SLWWWYYYYMMMWGGY	79.4
SRREEEERRRGGY	67.3	DRKEEEERRRRGGY	71.9	LKKELEERLRRLGGY	76.9	SALLELLLLLLLLGGY	79.4
SWEEEWRRRWGGY	66.1	SRREEEERRRRGGY	71.6	DELLELLLRLLLGGY	76.9	SLLELLLLLLLLLGGY	79.1
SRAEEEARRRGGY	65.7	SLAELLRRLLLGGY	71.4	DWWLEYWLRMWLGGY	76.8	EQRELEELLRRLLGGY	78.9
SEMEELLRRLGGY	65.7	SRLEEELRRRLGGY	71.4	SLWELMWLLMWLGGY	76.3	**DEERERLELLLRLGGY**	78.1
SALEELLRRLGGY	65.5	DRLRELLERLLGGY	71.3	DWRRYYEEMMRRGGY	76.2	SLLEELLRLLLLLGGY	77.7
SRAEEERRRRGGY	65.4	DRKLEEELRRRGGY	71.3	SRRELEELLRRLGGY	75.2	DAMRELLELLLRLGGY	77.7
DRRLEEELRRGGY	64.5	SWWEEYYRRMWGGY	71.0	DRLREELERRLRGGY	75.0	DLKRELEELLRRLGGY	77.6

[a] The amino- and carboxy-termini of the peptides were considered to be acetylated and amidated, respectively. In all cases, during the sequence optimization, the carboxy-terminal sequence of three amino acids was fixed at Gly-Gly-Tyr to simplify interpretation of the experimental measurement. Fractions of predicted α-helix were calculated with the current version of AGADIR. The peptides shown in bold were synthesized and its helix contents were measured by CD (see data of Table 1 and Figure 2).

helix stabilizing factors such as N- and C-capping motifs and electrostatic interactions with helix macrodipole and between the side chains. Certainly these trivial homopolymer motifs are not really useful due to its lack of solubility. However, there are many soluble sequences that are just a bit less stable than the homopolymer sequences.

In order to examine its helix stability, we synthesized a new series of four peptides (O13–O16) having one soluble representative for each peptide length. In Table 2 these peptides are marked in boldface. Figure 2 shows the CD spectra of this peptide series under the standard conditions (temperature 5 °C and pH 7.5) and their mean residue ellipticities at 222 nm and calculated helix contents are listed in Table 1. The CD spectra of all the peptides show typical α-helical shape with much lower negative CD signal at 208 and 222 nm as compared to that of the first series of synthetic peptides indicating further increase in its helix stability.

Figure 3 shows the correlation between the theoretical predictions obtained with current version of AGADIR model and the results of CD measurements for the synthetic peptides. One can see from the figure that there is reasonably high correlation between the theoretical predictions and the experimental measurements. Measurements done for O15 peptide at 10–100 μM concentrations showed that CD spectrum of this peptide has strong concentration dependence indicating influence of aggregation on the conformational stability of this peptide (data not shown). Therefore, O15 peptide was excluded from the correlation shown in Figure 3. This effect is probably due to the specific symmetrical sequence motif with interleaving pairs of hydrophobic (LL), negatively charged (EE), and positively charged (RR) amino acids promoting formation of a significant antiparallel β-structure fraction with favorable

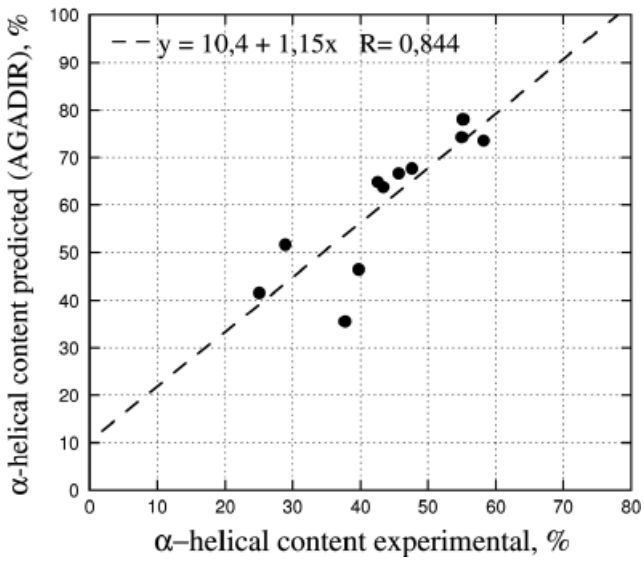

Figure 3. The correlation between the theoretical predictions and the experimentally determined helicities for peptides with optimized sequences.

interpeptide contacts between side chains of LL and LL as well as EE and RR. As for other peptides, only minor changes in the CD spectra were observed at high peptide concentrations. The results show that theoretical predictions tend to overestimate the α-helical content of these peptides (~10–20%), an indication that further modifications of the set of parameters are necessary for

correct predictions of helical contents for peptides with highly optimized sequences. However, even with the current parameterization, the method of global sequence optimization is capable to design peptide sequences with close-to-maximal number of stabilizing intrahelical interactions.

As discussed in the section Materials and Methods, GGY C-terminal sequence was introduced in the peptide sequences in order to simplify accurate measurements of peptide concentration only. This sequence certainly has strong negative impact on helix stability of peptides under investigation. It is of interest that AGADIR predictions of helix stability for the peptide series O13–O16 with truncated GGY C-terminal sequences are all around 90%. Taking into account the overestimations of helix content done by the current version of AGADIR, one can expect the maximum achievable helix stability in short peptides with fully optimized sequences at standard conditions to be ~70–75%.

Capping Box motifs [26] at the amino-terminus can be found in the somewhat less optimized peptide sequences indicating that in the absence of dominating hydrophobic side chain–side chain interactions, such motifs become an important factor in helix stabilization in agreement with previously reported data obtained for peptide and protein helices. Ser and Asp are the most frequent N-cap residues at these positions. One could expect that N1 and N2 (according to the nomenclature by Richardson and Richardson [27]) are occupied by negatively charged Glu and Asp utilizing stabilizing charge–helix macrodipole interactions and possible salt-bridges with a centrally disposed positively charged Arg. Despite a few such motifs found in Table 2, the more competitive solutions seem to be the hydrophobic interactions between Leu at N1, N2, and Leu at central positions of a peptide as well as series of centrally disposed negatively charged Glu and Asp forming salt bridges with positively charged Arg at both N- and C-termini of short peptides under considerations. In the latter case, of course, the peptides utilize the favorable interactions between a positively charged Arg residues and the helix macrodipole. There are also hydrophobic side chain–side chain interactions between residues at carboxy-terminal and central positions. In some cases, both these motifs are present simultaneously.

Despite its highest α-helical propensities, only very few peptide sequences have Ala residues. Although there are indeed a few cases, the number of centrally disposed salt-bridges in the optimized sequences is surprisingly low. The reason is probably associated with the influence of terminal positions in these very short peptides. It seems that hydrophobic residues (Leu) at central positions are more tolerant to the terminal requirements for accommodation of both negative charges from amino-termini and positive charges from carboxy-termini. Generally, the longer is a peptide, the more complicated are the patterns of sequential motifs that are found at the top of the list of best peptide sequences. The sequences in Table 2 utilize just a small fraction of protein amino acids that have high intrinsic helical propensities, and that can engage in side chain–side chain interactions and terminal capping. It seems these are the main factors playing a dominant role in the sequence space of globally optimized peptide helices.

It is widely accepted that sequences of protein helices are not, in general, optimized. The typical helical content of peptides derived from proteins is usually around 20% [10]. This may have certain biological 'sense' of keeping a balance between interactions within elements of secondary structure and those with the rest of a protein. This balance might act to prevent accumulation of nonnative intermediates during the folding of a protein. Therefore, we could not find a strong correlation between the optimized sequences presented in Table 2 and those in protein α-helices. However, motifs such as the Capping Box, interactions between charged residues and helix macrodipole, and hydrophobic interactions are often found in protein helices.

One of the important features of the proposed method is the possibility of arbitrarily fixing any functional segments of primary structure and optimizing just the nonfunctional elements. This option can be very useful in the case of helix optimization in globular proteins for the purposes of increasing its thermostability, for instance. In this case, only solvent-exposed amino acid positions of protein α-helices having local intrahelical contacts should be allowed to vary during the course of sequence optimization. These positions should be carefully selected by the analysis of protein 3D structure. All other amino acid positions of the helix should be fixed in their native sequence to preserve important tertiary interactions in the protein native structure.

In previous studies discussing possible reasons behind extreme thermostability of several proteins, one of us (M.P.) has reported that high stability of α-helices is a necessary condition for thermostability of several different protein families [28–30]. The magnitude of the observed decrease in intrinsic free energy upon α-helix formation of several thermostable proteins was found to be sufficient to explain the experimentally determined increase of their thermostability as compared to their mesophilic analogs. A natural reserve of RecA protein thermostability has been also estimated based on the sequences of RecA protein α-helices with close-to-optimal arrangement of solvent-exposed amino acids. It is surprising that the natural reserve of thermostability seems to be sufficient for enough conformational stability of the protein at nearly 200 °C.

Conclusions

In this study we have introduced a new practical method for the rational design of peptide and protein helices of the highest stability by the global optimization of primary structures. The method was found to be an efficient tool for peptide and, very likely, for protein engineering. Unlike the methods of global energy optimization (Molecular Dynamics, Monte-Carlo, Simulated Annealing, etc.) that are often used to change one or two amino acids in order to increase protein stability by adding a few favorable interactions to the structure of a protein, the method proposed here deals with all possible sequences of protein helices and selects the best one among them. To our knowledge this is first report of global sequence optimization reported in the literature to date.

Although, convergence to a global minimum at present cannot be rigorously proved for any interesting multidimensional problems, the method proposed here had an excellent efficiency in arbitrary fixed sequence optimization of peptide helices. Optimized sequences showed an expected increase of their helical content with peptide length. Sequence motifs found in the optimized sequences have also been reported to be major factors in the protein stabilization. CD measurements of helical content of peptides with optimized sequences indicate that maximum achievable helix content of 10–13 residue peptides with fully optimized sequences at 5 °C is expected to be as high as ~70–75%.

Acknowledgements

This work was supported by the Russian Foundation for Basic Research (Grant No. 07-04-00785-a), the Russian Academy of Sciences program 'Support to Innovations', by the Ministry of Education and Science of Russian Federation (grant RNP.2.2.1.1.4663), by St Petersburg government grant in the field of scientific and technological activities, by short-term fellowship from European Molecular Biology Organization, Germany (grant #ASTF 125-2005) and by short-term fellowship from the Agency of Industrial Science and Technology, Japan.

References

1 Huyghues-Despointes BM, Scholtz JM, Baldwin RL. Helical peptides with three pairs of Asp-Arg and Glu-Arg residues in different orientations and spacings. *Protein Sci.* 1993; **2**: 80–85.
2 Padmanabhan S, Baldwin RL. Tests for helix-stabilizing interactions between various nonpolar side chains in alanine-based peptides. *Protein Sci.* 1994; **3**: 1992–1997.
3 Armstrong KM, Fairman R, Baldwin RL. The (i, i + 4) Phe-His interaction studied in an alanine-based α-helix. *J. Mol. Biol.* 1993; **230**: 284–291.
4 Huyghues-Despointes BM, Scholtz JM, Baldwin RL. Effect of a single aspartate on helix stability at different positions in a neutral alanine-based peptide. *Protein Sci.* 1993; **2**: 1604–1611.
5 Lockhart DJ, Kim PS. Internal Stark effect measurement of the electric field at the amino terminus of an α-helix. *Science* 1992; **257**: 947–951.
6 Lockhart DJ, Kim PS. Electrostatic screening of charge and dipole interactions with the helix backbone. *Science* 1993; **260**: 198–202.
7 Aurora R, Rose GD. Helix capping. *Protein Sci.* 1998; **7**: 21–38.
8 Bryson JW, Betz SF, Lu HS, Suich DJ, Zhou HX, O'Neil KT, DeGrado WF. Protein design: a hierarchic approach. *Science* 1995; **270**: 935–941.
9 Villegas V, Viguera AR, Aviles FX, Serrano L. Stabilization of proteins by rational design of α-helix stability using helix/coil transition theory. *Fold. Des.* 1996; **1**: 29–34.
10 Munoz V, Serrano L. Elucidating the folding problem of helical peptides using empirical parameters. II. Helix macrodipole effects and rational modification of the helical content of natural peptides. *J. Mol. Biol.* 1995; **245**: 275–296.
11 Munoz V, Serrano L. Elucidating the folding problem of helical peptides using empirical parameters. III. Temperature and pH dependence. *J. Mol. Biol.* 1995; **245**: 297–308.
12 Lacroix E, Viguera AR, Serrano L. Elucidating the folding problem of α-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. *J. Mol. Biol.* 1998; **284**: 173–191.
13 Petukhov M, Munoz V, Yumoto N, Yoshikawa S, Serrano L. Position dependence of non-polar amino acid intrinsic helical propensities. *J. Mol. Biol.* 1998; **278**: 279–289.
14 Levy AV, Montalvo A. The tunneling algorithm for the global minimization of functions. *SIAM J. Sci. Stat. Comput.* 1985; **6**: 15–29.
15 Munoz V, Serrano L. Elucidating the folding problem of helical peptides using empirical parameters. *Nat. Struct. Biol.* 1994; **1**: 399–409.
16 Viguera AR, Serrano L. Experimental analysis of the Schellman motif. *J. Mol. Biol.* 1995; **251**: 150–160.
17 Petukhov M, Yumoto N, Murase S, Onmura R, Yoshikawa S. Factors that affect the stabilization of α-helices in short peptides by a capping box. *Biochemistry* 1996; **35**: 387–397.
18 Petukhov M, Uegaki K, Yumoto N, Yoshikawa S, Serrano L. Position dependence of amino acid intrinsic helical propensities II: non-charged polar residues: Ser, Thr, Asn, and Gln. *Protein Sci.* 1999; **8**: 2144–2150.
19 Petukhov M, Uegaki K, Yumoto N, Serrano L. Amino acid intrinsic α-helical propensities III: positional dependence at several positions of C terminus. *Protein Sci.* 2002; **11**: 766–777.
20 Himmelblau DM. *Applied Nonlinear Programming*. McGraw-Hill: New York, 1972.
21 Miele A, Huang HY, Heideman JC. Sequential gradient-restoration algorithm for the minimization of constrained functions – Ordinary and conjugate gradient versions. *J. Optim. Theor. Appl.* 1969; **4**: 213–243.
22 Bartenev OV. *FORTRAN for Professionals*, Vol. 1. Dialog-MIPI: Moscow; 2000.
23 Chen YH, Yang JT, Chau KH. Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. *Biochemistry* 1974; **13**: 3350–3359.
24 Chakrabartty A, Kortemme T, Padmanabhan S, Baldwin RL. Aromatic side-chain contribution to far-ultraviolet circular dichroism of helical peptides and its effect on measurement of helix propensities. *Biochemistry* 1993; **32**: 5560–5565.
25 Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 1983; **22**: 2577–2637.
26 Harper ET, Rose GD. Helix stop signals in proteins and peptides: the capping box. *Biochemistry* 1993; **32**: 7605–7609.
27 Richardson JS, Richardson DC. Amino acid preferences for specific locations at the ends of α-helices. *Science* 1988; **240**: 1648–1652.
28 Petukhov M, Kil Y, Kuramitsu S, Lanzov V. Insights into thermal resistance of proteins from the intrinsic stability of their α-helices. *Proteins* 1997; **29**: 309–320.
29 Petukhov MG, Kil' Iu V, Lantsov VA. Nature of protein thermal resistance: stability of α-helices of RecA proteins of the thermophilic bacteria. *Dokl. Akad. Nauk* 1997; **356**: 268–271.
30 Petukhov MG, Baitin DM, Kil' Iu V, Lantsov VA. Optimal rigidity of the protein structure: three classes of rigidity in the eubacterial RecA protein family. *Dokl. Akad. Nauk* 1998; **362**: 118–121.

