Use of molecular mechanics for secondary structure prediction. Is it possible to reveal α-helix?

Gelena T. Kilosanidze[a], Alexey S. Kutsenko[b], Natalia G. Esipova[a], Vladimir G. Tumanyan[a,*]

[a] *Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow 119991, Russia*
[b] *Centre for Genomics and Bioinformatics, Karolinska Institute, P.O. Box 280, S-171 77 Stockholm, Sweden*

Received 3 October 2001; revised 20 November 2001; accepted 20 November 2001

First published online 30 November 2001

Edited by Gunnar von Heijne

Abstract A new approach to predicting protein standard conformations is suggested. The idea consists in modeling by molecular mechanics tools a continuous α-helical conformation for the whole protein. The profile of energy along the model α-helix reveals minima corresponding to real α-helical segments in the native protein. The 3/10-helices and β-turns including a local α-helical conformation may be detected as well. All α-helical segments in the test sample are delineated; mean residue by residue accuracy $Q_{3\alpha}$ is 79%. This non-statistical approach can shed light on the physical grounds of α-helix formation.

Key words: Protein secondary structure; Structure prediction; α-Helix; Molecular mechanics

1. Introduction

In spite of the progress in protein secondary structure prediction, alternative approaches appear to be necessary for understanding the physical grounds of structure formation. We try to demonstrate that conventional conformation analysis may be employed for this purpose.

It is reasonable to treat the set of conformations in a native protein as a superposition of typical standard conformations. Among them we know α-helix, β-structure, and left-handed helix of polyproline II type. The localization of a certain conformation in a definite site of polypeptide chain must to be consistent with local and long-range interactions and with the global minimum of energy for the whole protein. A common way to do the conformation search by the molecular mechanics method implies generation of a combinatorial set of conformations. After this and subsequent energy minimization, the conformation with favorable energy may be treated as realistic. However, for standard conformations having obvious advantages in energy, e.g. α-helix, a combinatorial choice of conformations seems superfluous.

The idea of the work may be outlined as follows. If we formally assume one and the same type of conformation for the entire protein chain, the segments that actually have this conformation in the real protein will get advantages in energy. The α-helical conformation is naturally the first candidate for such study, since α-helix is a tight structure simultaneously stabilized by hydrogen bonding, and Coulomb and van der Waals energies.

Physical conditions as well as local and long-range interactions play an important role in secondary structure formation [1–4]. Nevertheless, contemporary methods for secondary structure prediction are based on statistics and even on direct resemblance between the primary structures within a protein family [5–7].

Molecular mechanics is employed for protein conformation analysis in general. In particular, it may provide a new independent view on the problem of secondary structure. We hope this approach will serve to explain the sequence effect and the role of various energy terms in the establishment of α-helical conformation.

2. Material and methods

2.1. Protein sample

Eight proteins were taken from PDB database with account for the structural class according to the CATH classification (http://www.biochem.ucl.ac.uk/bsm/cath) [8]. The choice was random within the following conditions: resolution <2.5 Å, no disulfide bridges, no ligands or ions interacting with α-helical segments.

2.2. Modeling of α-helical conformation

By ICM molecular mechanics program [9,10] using ECEPP/3 parameters [11], all protein chains were folded into a continuous α-helix; starting values of dihedral angles were $\phi=-57°$, $\Psi=-47°$. Optimal conformations of side groups were found by a standard Monte Carlo procedure using limited grid search of conformations. The number of variations of variables in average exceeds 1000 for all dihedral angles of a side chain. The starting temperature was 600 K. For explicit description of the optimization procedure, see [10]. At the last step of optimization, the ϕ,Ψ angles of the backbone were restrained. The penalty function switched on beyond $\pm 10°$ around ideal α-helical ϕ,Ψ values.

2.3. Modeling of extended baseline conformation

To take into account constant and nearly constant contributions in energy that vary from residue to residue, it is necessary to subtract the energy of an unrestrained conformation. The baseline conformation was obtained by altering the ϕ,Ψ angles of model α-helix conformation to $-160°$,$+160°$ respectively as in [12].

With the aim of removing some possible hindrances in the extended conformation, additional optimization was performed to adjust the χ angles in side chains and ϕ,Ψ angles in the main chain. This optimization practically does not alter the conformation.

2.4. Energy profile construction

The energies of pentapeptides (or any oligopeptides) were estimated

*Corresponding author. Fax: (7)-095-1351405.
E-mail address: tuman@imb.imb.ac.ru (V.G. Tumanyan).

0014-5793/02/$22.00
PII: S0014-5793(01)03212-4

separately for the helical and for the extended conformation. The resultant profile is a subtraction of energies of these conformations.

2.5. Residue by residue performance

We followed the standard $Q_{3\alpha}$ formula for two-state prediction (positive correct and negative correct predictions) as in [13].

3. Results and discussion

Figure 1 shows coincidence of minima on the continuous α-helix energy profiles and the real α-helices for the second domain of the DNA-binding protein (1pdn). Regions of β-structure, coils, and some β-turns correspond to high energy levels and barriers in the profile. Fig. 1 illustrates also the procedure of energy profile construction by subtraction of two curves. It is obvious that energy variations for the extended conformation have minor amplitudes. Thus, the curve for model α-helix alone bears essentially all meaningful peculiarities. Note (Fig. 1C) that the van der Waals component of energy reproduces in general the trace of total energy. We may conclude that even for the best side groups conformations, the sequence in linking regions between real α-helical segments remains less favorable for α-helix formation.

Another example of the protein from the all-α class is shown in Fig. 2. Both α-helices of the Arc receptor (1baz) may be reproduced by the long α-helix modeling procedure. Additionally, Fig. 2 shows that the method is practically unaffected by fragment length chosen for local energy estimation. The choice of a pentapeptide has some advantages. Indeed, the first hydrogen bond forms at this length. Using shorter fragments leads to some noise, and longer ones smooth the borders of the low-energy regions, as can be seen for tripeptide and heptapeptide fragments (Fig. 2B,C). Although very short fragments provide better resolution, it is difficult to take a universal baseline for various proteins in this case. The pattern of energy profiles is apparently retained for various force fields. The positions of maxima and minima are conserved for 1baz and other proteins after including, as variants of the surface term, the atomic solvation

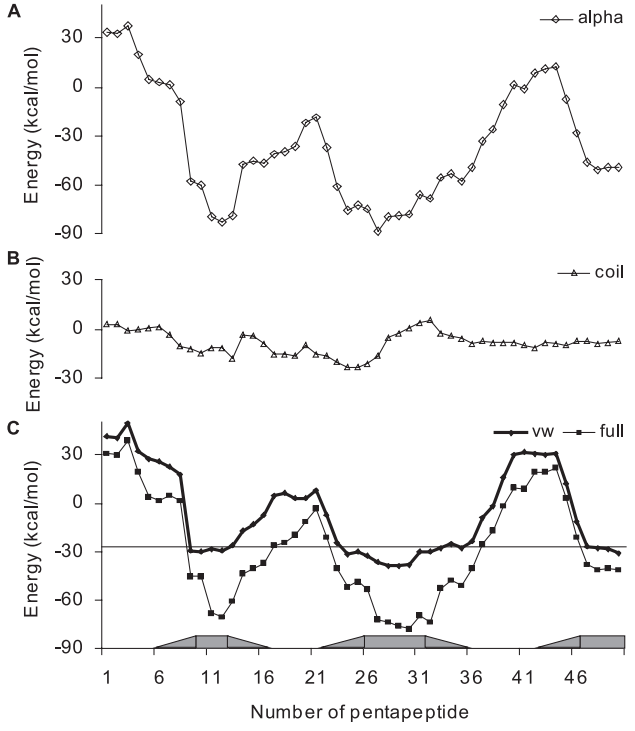

Fig. 1. Energy profiles for the second domain of regulatory DNA-binding protein (1pdn). (A) Energy of modeling α-helical conformation; (B) energy of baseline extended conformation; (C) resulting full energy after subtracting (B) from (C) and van der Waals component (VW) of resulting full energy. Each point *i* corresponds to the energy calculated for a pentapeptide comprising residues *i* to *i*+4. The first residue of the first pentapeptide has number 71 according to residue numeration in the protein chain. Rectangle and adjunct triangles denote α-helical pentapeptides and pentapeptides including at least one α-helical residue, respectively, in accordance with the X-ray data. The line for prediction of α-helices is drawn at −28 kcal/mol.

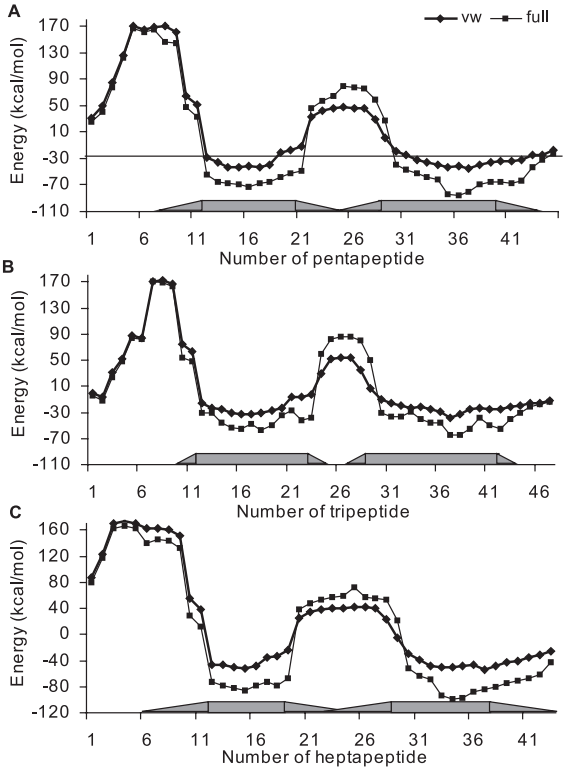

Fig. 2. Profiles of full energy and van der Waals component for Arc receptor (1baz) for various lengths of the fragment. (A) Penta-, (B) tri-, and (C) heptapeptide. All notations as in Fig. 1. The first residue of the first oligopeptide has number 5 according to residue numeration in the protein chain.

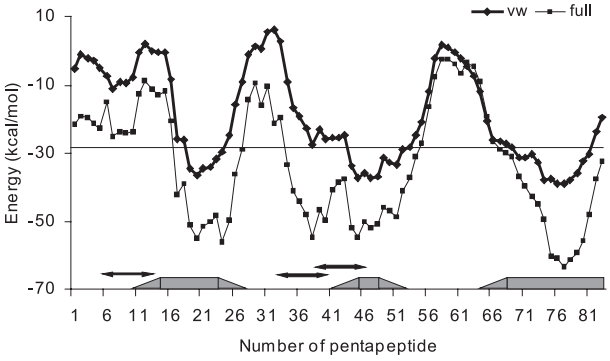

Fig. 3. Profiles of full energy and van der Waals component for phosphotransferase (1ptf). All notations as in Fig. 1. The first residue of the first pentapeptide has number 1 according to residue numeration in the protein chain. Additionally, pentapeptides including β-turns are marked by arrows in accord with X-ray data.

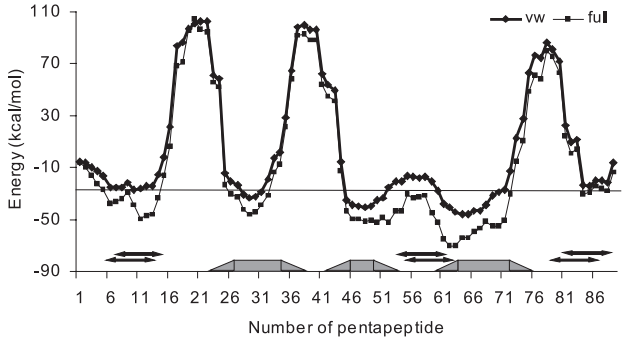

Fig. 4. Profiles of full energy and van der Waals component for the transcription regulation domain of 1opc protein. Notations as in Figs. 1 and 3. The first residue of the first pentapeptide has number 137 according to residue numeration in the protein chain.

or the polarization added to the Coulomb term for electrostatic energy evaluation following standard procedures in ICM [9,10] (data not shown).

Interestingly, the method is sensitive to even short segments of α-helical conformation. For histidine-containing phosphocarrier protein (1ptf), as Fig. 3 shows, there are minima which may be attributed to the three β-turns (all type I). Note that in β-turns of I types there are two consecutive residues with α-helical ϕ,Ψ values. These minima are narrow and lie higher the prominent main minima corresponding to the α-helices in the protein. The van der Waals curve is most useful to distinguish obvious minima for α-helices and rather subtle minima for β-turns. On the van der Waals curve, all β-turn minima lie above the α-helix ones.

Fig. 4 for the transcription regulation domain of protein 1opc exemplifies, above all, the effect of proline on the shape of the profile. There are some shortcomings in Pro modeling in most molecular mechanics programs. The Pro ring is frozen in a rigid conformation, which leads to overestimating or underestimating the energy of the Pro-containing fragments to an extent that depends on the concrete conformation; overestimation is more frequent in practice (see Fig. 4). Prolines are at position 158, 160, 176, 179, 216 and 219. To take into account the Pro influence rigorously, it is desirable to introduce a flexible pyrrolidine ring. Regarding the conformational programs in general, a flexible Pro moiety would obviously be useful to maximize the reliability of prediction, because the borders of energy wells can be estimated more precisely.

Table 1 summarizes the data for whole protein sample. Let us discuss the prediction power of the method suggested. Firstly, the van der Waals component is most convenient in this respect as compared with electrostatic and H-bonding terms. The minima for α-helices are readily observed. Secondly, the mean value of these minima seems to be constant. We found that the −28 kcal/mol cutoff is most suitable for prediction. Only one helical element, namely the terminal 3/10-helix, has not been reproduced (see eighth row in Table 1). Generally, helices of the 3/10-type may be traced, the mean energy for 3/10-helices is slightly above the α-helical mean energy. At the same time, for replication terminator protein (see second row in Table 1) we have one superfluous helical fragment. That is, the underpredictions and overpredictions of the helical segments are few. Standard per-residue criteria (see Table 1) testify that the performance of the method is quite good. The mean values for $Q_{3\alpha}$, Q_{α}, and $Q_{\text{non-}\alpha}$ are 79, 74.4, and 84.4%, respectively. As for predictions of β-bends, particularly those that include local α-helical conformation, they are sometimes difficult to reproduce, and the mean value for $Q_{\beta-\text{turn}}$ does not exceed 45% (the baseline was set at −19 kcal/mol in this case). The cases of underpredictions in the second and fourth rows of Table 1 are connected with the proline effect. Note that the method was not intended for precise prediction of very short segments of α-helical conformation. This is an unexpected possibility of the approach proposed.

Thus, the segments of polypeptide chain that are predisposed to α-helix formation and are helical in native protein, retain this property in continuous modeling of α-helix for the whole chain. In this way, prediction of α-helical conformation including α-helices and even β-bends with appropriate ϕ,Ψ values becomes quite possible. We can also delineate the regions that can never adopt α-helical conformation. Simultaneously, the physical grounds of α-helix formation become more understandable. In particular, the main contribution to energy is the van der Waals component. This component is sufficient for domain protein structure determination [14]. The van der Waals energy is also sufficient for choosing the correct fold for a particular primary structure [15]. Thus, the role of local interactions in establishing a definite type of secondary structure can be traced, also in accordance with [2,3]. The results obtained suggest that the influence of tertiary structure is not so significant. The computational protocol described here uses readily available molecular mechanics pro-

Table 1
Summary of the computation results

No.	PDB code	Protein	Class	Residue No.	α-Helices		Per-residue accuracy[a] (%)			β-Turns (α-helical conf.)	
					X-ray	Predicted	$Q_{3\alpha}$	Q_{α}	$Q_{\text{non-}\alpha}$	X-ray	Predicted
1	1baz	Arc repressor	α	5–53	2	2	73	61	94	0	0
2	1bm9	Replication terminator protein	α	3–122	3	4	76	81	73	3	2
3	1igd	Immunoglobulin-binding protein G	α+β	1–61	1	1	88	100	84	3	3
4	1opc	OmpR (transcription regulation)	α	137–229	3	3	84	61	97	8	6
5	1pdn	Paired protein	α	71–124	3	3	84	65	100	0	0
6	1pdo	Phosphoenolpyruvate-dependent phosphotransferase	α/β	2–130	5	5	73	78	67	2	3
7	1ptf	Histidine containing phosphocarrier protein	α/β	1–87	3	3	86	74	94	3	3
8	2chs	Chorismate mutase (isomerase)	α/β	2–115	5[b]	4	68	75	66	0	0

[a]Residue by residue accuracy of secondary structure prediction: $Q_{3\alpha}$ − α and non-α, Q_{α} − only α, $Q_{\text{non-}\alpha}$ − only non-α.
[b]Two α-helices and three 3/10-helices

grams and, albeit comparatively time consuming, offers a way of a priori secondary structure analysis and prediction.

Acknowledgements: The authors are grateful to A. Galkin for critical reading of the manuscript and valuable comments. This work was supported in part by the Russian Foundation for Basic Research (Grants 00-04-48351, 00-07-90037 and 00-15-97834) and INTAS 99-1476.

References

[1] Lim, V.I. (1975) J. Mol. Biol. 88, 873–894.
[2] Ptitsyn, O.B. and Finkelstein, A.V. (1983) Biopolymers 22, 15–25.
[3] Ptitsyn, O.B. and Finkelstein, A.V. (1989) Protein Eng. 2, 443–447.
[4] Kuwajima, K., Semisotnov, G.V., Finkelstein, A.V., Sugai, S. and Ptitsyn, O.B. (1993) FEBS Lett. 334, 265–268.
[5] Persson, B. (2000) EXS 88, 215–231.
[6] Baldi, P., Brunak, S., Chauvin, Y., Andersen, C.A. and Nielsen, H. (2000) Bioinformatics 16, 412–424.
[7] Zhang, C.T. and Zhang, R. (2001) Proteins 43, 520–522.
[8] Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B. and Thornton, J.M. (1997) Structure 5, 1093–1108.
[9] Abagyan, R., Totrov, M. and Kuznetsov, D. (1994) J. Comp. Chem. 15, 488–506.
[10] Abagyan, R. and Totrov, M. (1999) J. Comp. Phys. 151, 402–421.
[11] Nemethy, G., Gibson, K.D., Palmer, K.A., Yoon, C.N., Paterlini, G., Zagari, A., Rumsey, S. and Scheraga, H.A. (1992) J. Phys. Chem. 96, 6472–6484.
[12] Pitera, J.W. and Kollman, P.A. (2000) Proteins 41, 385–397.
[13] Schulz, G.E., Schirmer, R.H., 1979, Principles of Protein Structure, pp. 1–314, Springer Verlag, New York.
[14] Berezovsky, I.N., Tumanyan, V.G. and Esipova, N.G. (1997) FEBS Lett. 418, 43–46.
[15] Petrey, D. and Honig, B. (2000) Protein Sci. 9, 2181–2191.