Journal of Bioinformatics and Research
Vol. 1, No. 4 (2012): 41-45
Research Article
Open Access

aizeon publishers

ISSN: 2277-3363

Tips for loop prediction and conformational search

Jamal Raiyn, Mohammed Azab, Mahmud Masalha and Anwar Rayan*

Drug Discovery Informatics Lab, QRC-Qasemi Research Center,Al-Qasemi Academic College, P.O.B. 124, Baka El-Garbiah 30100, Israel
* Corresponding author. Tel: +972-4-6286761/4; email: a_rayan@qsm.ac.il

Received: 03 September 2012 Accepted: 21 September 2012 Online: 05 October 2012

ABSTRACT

The 3D structure determination of a protein is crucial for structure-guided drug development and the homology modeling approach is the most accurate method among the computational methods, yielding reliable models. Loop predictions are required in many protein homology modeling studies and raise the problem of loop closure. Loop closure problems have been solved mainly by "ab-initio" methods or by employing databases. "ab-initio" methods mostly use standard bond lengths and angles. However, backbone bonds and angles are not standard in different crystallographic structures, thus, is it possible to achieve high quality loop closure by employing standard bond lengths and angles? To explore this issue we reconstructed loops from the structurally refined proteins with standard parameters but retained the experimental backbone dihedral angles. This was tested by stepwise construction from one of the terminals (N- and C-) toward the other as well as from both terminals towards the center. We conclude that introducing variability for bond lengths, for bond angles and for omega dihedrals, is essential for accurate prediction of loops. As well, constructing loops from both terminals toward the center give more precise models, closer to native structures.

Keywords: homology modeling, loop closure, optimization

INTRODUCTION

The 3D structure determination of a protein is crucial for structure-guided drug discovery and development [1-4]. Since experimental structures are available only for a small number of sequenced proteins [5], alternative strategies are required to predict reliable models for proteins whose structures not determined yet by X-ray diffraction or NMR [6].

Homology Modeling (HM) also termed Comparative Modeling (CM) is useful if a sequence has ≥ 30% identity to another sequence with known structure [7, 8]. In HM techniques, pair-wise or multiple sequence alignments suggest which parts of the sequence may be "mapped" into the known structure. In many cases, helices and sheets are found to be structurally conserved regions (SCRs), while the structurally variable regions are the protein "loops" or "coils", those stretches of sequence whose structure can not be defined by standard geometric parameters and are "connecting" secondary structure elements. Thus, the main predictive efforts in homology/comparative modeling are directed to these variable loops and to the side chains [9-11].

Loop modeling may be handled by both "ab-initio" methods or by database search techniques. "ab-initio" loop prediction is based on a conformational search and enumeration of conformations in a given environment, guided by a scoring function [12], while database approaches to loop prediction consist of finding segments of main chain from a database, so that they fit the two stems of the desired loop [13].

Modeling a loop requires to satisfy the constraints of connecting the two protein segments on either end of the loop with a physically reasonable peptide conformation. The stem regions of a loop set constraints on the available conformations, thus reducing the size of the conformational space, but satisfying these constraints still presents an algorithmic challenge.

In "ab-initio" methods, if a loop is constructed from the N-terminal anchor, then the loop must be adjusted to properly connect the protein at the C-terminal anchor. If the loop is constructed from both the N- and C- terminal anchors, the resulting segments must fit and be connected at the middle of the loop. Two main issues in loop prediction are the accuracy (measured by root mean square deviation "RMSD") and the loop length. These two

are not necessarily disconnected, as it is expected that predicting longer loops might be less accurate.

Many "ab-initio" loop prediction methods, employ standard bond lengths and angles [14-16] or have demonstrated their technique with the original bond lengths and angles while optimizing only the backbone dihedrals [17], and thus have to deal with the "loop closure" issue. It is important to remember that "success" in loop prediction is measured against known protein structures. A successful loop closure while keeping its experimental bond lengths and angles may not necessarily be valuable in homology modeling, where bonds and angles are unknown.

Several solutions have been presented to solve the "loop closure" problem with standard bonds, bond angles and peptide dihedrals: Go and Scheraga were the first to develop a procedure for predicting the conformation of a fragment joining two polypeptides of known structure. The Go and Scheraga algorithm [18] finds analytical solutions to the ϕ and ψ angles for a stretch of 3 residues, given the position and orientation of each end. The work of Coutsias et al [15] is a modification of the Go and Scheraga procedure that allow small bond angle variations.

Moult and James [19] generated different conformations by using the dihedral angles around single bonds in a polypeptide backbone (ϕ and ψ) and side chains (χ_n) with standard geometry for the bond lengths, bond angles, and peptide torsion angles. Loops were constructed from both N- and C- anchors, toward the center, with a set of main-chain conformations for each residue of these half lengths. The geometry of each complete main-chain conformation (formed by up to 11 pairs of ϕ and Ψ dihedral angles for each amino acid) is then adjusted by energy minimization to obtain bond lengths, bond angles, and peptide dihedrals throughout that are within acceptable deviations from the standard values.

Shenkin et al [20] devised the "random tweak" algorithm. Starting from a random conformation, all dihedral angles are modified at once in each step of the iteration until the distance constraints between the end residues are satisfied. The algorithm involves the variation of only dihedral angles. Ensembles of 100 properly closed backbone structures for each loop were generated under several conditions of Van-der-Waals interactions within the loop and with the rest of the protein. These authors tested the method on 6 CDR loops of an immunoglobulin, with lengths of 5-19 residues. Xiang et al [21] extended the use of the random tweak to form "colonies" of loops based on energies and RMSD values, in an iterative process, and obtained good results. In the "colony energy, which includes a measure of the entropy, the average global RMSD of 8- and 12-residue loops is 1.45Å and 3.42Å, respectively.

Sudarsanam et al [22] described a loop modeling procedure which uses a database of $\phi_{(i+1)}$, $\Psi_{(i)}$ backbone values from some 50,000 dimers in a non-redundant version of the PDB. Loops were predicted by taking backbone dihedral angle values for the appropriate dimer in the sequence from the database, and generating 10,000 to 50,000 loop conformations, depending on the loop length. The resulting fragments, of length 5-9 residues, were then filtered using the geometric restraints of the flanking residues.

Deane and Blundell [14] developed an approach for the "*ab-initio*" generation of an exhaustive set of candidate fragments, which are stored in a database that can be used for modeling any region of a polypeptide. These candidate fragments are then evaluated using a set of rule-based filters to find the best polypeptide fragment in the environment of the target protein. The polypeptide fragments are from a computer-generated database encoding all possible peptide fragments up to twelve amino acid long. Each amino acid can be introduced with one out of eight possible Φ/ψ pairs, that were obtained in order to represent the protein databank. DePristo et al [16] suggested a method that samples fine-grained, residue-specific Φ/Ψ state sets to find conformations that satisfy a number of geometric and knowledge-based filters, including reasonable Φ/Ψ angles, gap-closure and excluded-volume restraints. Applied to a large number of loop structures, this method samples consistently near-native conformations, averaging 0.4, 1.1 and 2.2Å for main-chain RMSDs of 4, 8 and 12 residue long loops.

The CCD algorithm of Canutescu and Dunbrack [17] employs experimental angles and adjusts one dihedral angle at a time to minimize the sum of the squared distances between three backbone atoms of a moving C-terminal and the corresponding atoms in the fixed C-terminal. The algorithm proceeds iteratively through all of the adjustable dihedral angles from the N-terminal to the C-terminal end of the loop. In the CCD algorithm, the average of the best backbone RMSD for loops of 4, 8 and 12 amino acids is 0.56, 1.59 and 3.04Å, respectively. The CSJD algorithm of Coutsias et al [15] generalizes previous work on analytical loop closure and allows for a small degree of flexibility in the bond angles and the peptide torsion angles. For constructing larger loops the authors employ an existing loop construction algorithm [23] which samples the allowed regions of the Φ/Ψ map in a discrete manner while screening possible side chain clashes with a rotamer library. This algorithm is used to construct the N- and C-terminal branches except for a three residue gap in the middle of the loop, and the analytical loop-closure algorithm is used to close that gap. The best RMSD values obtained from this method are 0.40, 1.01, and 2.34Å for 4, 8 and 12 residue loops, respectively.

It is thus common to employ dihedral angles ϕ and ψ from experimental results or by rotating them incrementally, to achieve loop closure. This is more tedious with larger loops. Dihedral angle variations are detected more easily

by the "naked eye" of the researcher, while much less attention is directed to variations of bond lengths and angles. This is also the case of peptide dihedral angles, which are easily characterized as "cis" or "trans". Considering the variations of bond lengths, angles and peptide dihedrals is computationally more demanding. Therefore, it is more common to use standard values for these three types. The main question that we pose here is: what is the effect of using such standard values on the success in loop prediction? This is a relevant question in most of the cases mentioned above: it is intricately related to the limit on the number of variables that one introduces into a general method for loop closure, and it is thus some function of the loops' length.

In this paper we demonstrate the influence of standard bond lengths on loop closure of several loop sizes and evaluate the deviations, caused by using standard values, from experimentally determined conformations.

METHODS

Test Set

Twenty eight loops were studied, of which ten were tetramers, ten octamers and eight 12-mer loops. Construction of these loops by several methods was compared (*vide infra*). They are parts of structures that have been solved to a resolution better than 1.6Å and have a low mutual sequence identity < 20%.

Reconstructing experimentally determined loops

Experimentally determined loops were "cut off" from the protein and reconstructed with standard values for bond lengths, bond angles and ω dihedral angles, while retaining the experimentally determined ϕ and Ψ dihedral angles. Standard values employed for bond lengths and angles, from the AMBER potential energy function, are presented in table 1. In table 1, we also present the averages and standard deviations for these bonds and angles in our own database of twenty eight loops and in the literature. The two stems at the N- and C- terminals of the loops were retained for the loop positioning, while the rest of the protein structure played no role in this examination.

Table 1. Averages and standard deviations of standard bond length, bond angles and omega dihedral angles: experimental parameters, measured values of the twenty eight analyzed loops, and values used as standard values in the analysis.

		Mean[A]	Std. Dev.[A]	Mean[C]	Std. Dev[C]	values[D]
Bond length [Å]	C-N	1.329	0.014	1.329	0.010	1.335
	N-CA	1.458	0.019	1.458	0.012	1.449
	CA-C	1.525	0.021	1.522	0.010	1.522
Bond angle [^{0}]	CA-C-N	116.2	2.0	117.0	2.0	116.6
	C-N-CA	121.7	1.8	121.5	2.4	121.9
	N-CA-C	111.2	2.8	111.7	3.4	110.3
Dihedral Angle[^{0}]	ω	180.0[B]	5.8[B]	179.9	4.9	180.0

[A] Values taken from Engh &Huber [24]. It is worth to assign that the Engh and Huber values are used as stereo-chemical restraint targets in most macromolecular refinement programs [25].
[B] Values taken from Morris et al [26].
[C] Values calculated over the 28 loops considered in our analysis.
[D] Standard values used in our analysis taken from the AMBER potential energy function [27].

Construction was achieved by two methods. In the first, loops were constructed stepwise from one of the terminals (N- or C-) toward the other, and in the second, construction took place from both terminals concurrently, to their meeting point at the center. The influence of ω dihedral angles was examined by repeating the construction with both these methods, while retaining the experimental ω values.

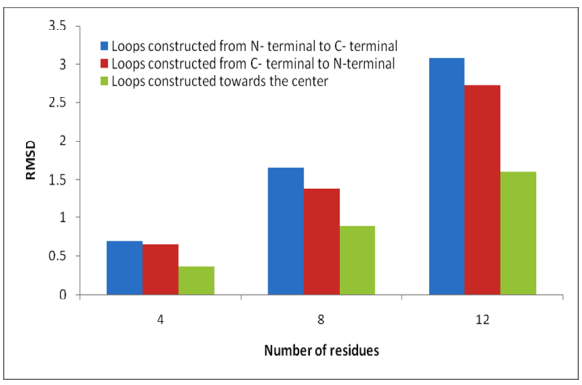
RESULTS AND DISCUSSION

In this work we examine how the use of standard bond lengths and bond angle affect the success in loop predictions.

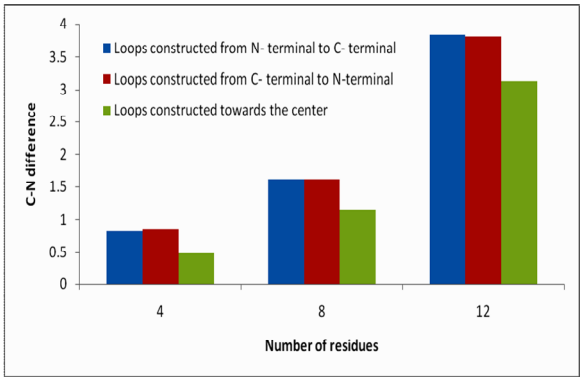
Reconstructing loops with standard values

We reconstructed experimentally determined loops by replacing their bond lengths and bond angles, as well as their peptide dihedral angles with standard values, while keeping the crystallographically determined ϕ and ψ angles all along. With these values, we compared three stepwise loop construction techniques in search for a preferred one, in which the errors introduced by standard values may hopefully be kept to a minimum.

Constructing loops from N- to C- terminals, with standard bond lengths, bond angles and ω dihedral angles resulted in deviations from experiment, that are shown in figure 2, together with the results of loop construction in the opposite, i.e., C- to N- direction as well as by simultaneous construction from both terminals toward the center. Figure 2a presents the RMSD values of the constructed loops compared to experimental, in a histogram, for the various loop sizes. As well, figure 2b presents the average absolute differences between the standard C-N bond length and the distance between C and N atoms at the meeting point.

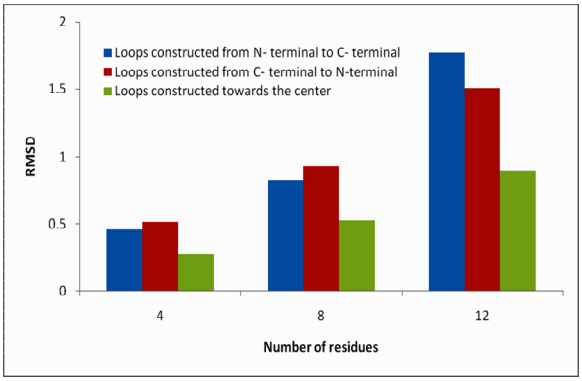


2a. Average RMSD for the various constructed loops, in comparison to experiment.

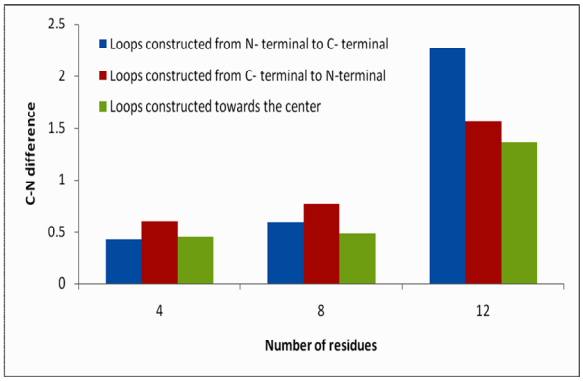


2b. Average absolute differences between the standard C-N bond length and the distance between C and N atoms at the meeting point

Figure 2. Deviations from experiment while reconstructing experimentally determined loops with standard bond lengths, standard angles and standard ω dihedral angles.



3a. average RMSD for the various constructed loops, in comparison to experiment.



3b. average absolute differences between the standard C-N bond length and the distance between C and N atoms at the meeting point

Figure 3. Deviations from experiment while reconstructing experimentally determined loops with standard bond lengths, standard angles and experimentally determined ω dihedral angles.

Building the loops from N- to C- terminals, with standard bond lengths and bond angles, but retaining each loop's experimentally determined ω dihedral angles, resulted in much smaller deviations from experiment, that are given in figure 3, together with the results of loop construction in the opposite direction and simultaneous construction toward the center. RMSD values for loop construction with experimental ω are presented in figure 3a. Average absolute differences between the standard C-N bond length and the distance between C and N atoms at the meeting point, are shown in figure 3b.

Loops constructed from both terminals simultaneously have significantly lower RMSD values in comparison to loops that are constructed from one terminal toward the other. As could be expected, there is no apparent difference between RMSD values for loop construction from N- to C- terminal or in the opposite direction. It is possible that, when the loop is built from both ends, there is a better dispersion of accumulated errors. Thus, construction of loops in loop prediction methods from both terminals toward the center seems to be a better practice.

In the stepwise addition of residues by all three methods, RMSD values become worse as the number of residues in the reconstructed loops increases. Retaining the experimentally determined ω dihedral angles decreases the average RMSD values to ~50-80% of their original size. It is not clear to us why the introduction of correct ω values has such a dramatic effect. Part of the reason could be the larger standard deviation (Table 1) of these dihedrals compared to the deviations of bonds and angles. However, the substantial improvement by using experimental ω dihedrals rather than the standard 180° ones has no easy practical outcome.

"*ab initio*" prediction of loop structure by stepwise construction can not be based on precise bond lengths and angles, because they are unknown. The alternative strategies are: 1. Use standard bonds and angles [14-16]; 2. Apply variations for bond lengths and bond angles, as well as for the peptide dihedral angle and 3. Use amino acid units from experimental results [13, 17]. In the second strategy, there could be many options for each unit rather than a single "standard" which has average bond lengths and angles. As a result, the problem of loop construction becomes "combinatorial" in nature. Therefore, this second approach requires a stochastic construction of many alternatives [13], with a reliable evaluation that includes the loop closure requirement. The third approach calls for generating the equivalents of "Ramachandran plots" for bond lengths and angles of amino acids, or just to vary each by some reasonable amount within specific limits, which could be derived from table 1. This would include also variations in the ω angles. Such an approach requires, again, a method that can deal with the combinatorial nature of the many options that are produced. Construction according to the first option, with standard values, requires a combinatorial solution as well [12]. In our study we reconstructed loops by keeping their original dihedral angles. But, in "real life" predictions, the dihedrals must be searched as well. Thus, it is again a combination of the angles along the backbone that will determine the loop closing ability. In that case, we suggest to construct from both terminals simultaneously, toward the center [13]. This approach minimizes the deviations, whatever the length of the loop.

Finally, some of the reports in the literature of loops and of cyclic peptides ("self closing loops") present only dihedral angles and not bond lengths and bond angles. Attempts to reconstruct such loops from the dihedral angles only fail, due to the importance of the bond length and angle variations between residues. It is thus inappropriate to assume that reporting dihedral angles alone may be sufficient for loop closures.

CONCLUSIONS

The inherent flexibility of amino acid parameters should be taken into account in any high-resolution structure prediction in proteins. Two tips that could be useful while predicting loop structure: 1) introducing variability for bond lengths, bond angles and for ω dihedrals is essential for accurate prediction of loops; 2) loop construction from both terminal towards the center is preferable, for minimizing the errors caused by using standard values.

Acknowledgments

The authors would like to thank Al-Qasemi foundation for supporting publication of this manuscript.

REFERENCES

1 Patny A, Desai PV and Avery MA. (2006) Homology modeling of G-protein-coupled receptors and implications in drug design. Curr Med Chem **13**, 1667-1691

2 Ferrara P and Jacoby E. (2007) Evaluation of the utility of homology models in high throughput docking. J Mol Model **13**, 897-905

3 Rajamohan F, Marr E, Reyes AR et al. (2011) Structure-guided inhibitor design for human acetyl-coenzyme A carboxylase by interspecies active site conversion. J Biol Chem **286**, 41510-41519

4 Han Z, Pinkner JS, Ford B et al. (2010) Structure-based drug design and optimization of mannoside bacterial FimH antagonists. J Med Chem **53**, 4779-4792

5 Neumann S, Hartmann H, Martin-Galiano AJ et al. (2012) Camps 2.0: exploring the sequence and structure space of prokaryotic, eukaryotic, and viral membrane proteins. Proteins **80**, 839-857

6 Hazai E and Bikadi Z. (2008) Homology modeling of breast cancer resistance protein (ABCG2). J Struct Biol **162**, 63-74

7 Rayan A. (2009) New tips for structure prediction by comparative modeling. Bioinformation **3**, 263-267

8 Baker D and Sali A. (2001) Protein structure prediction and structural genomics. Science **294**, 93-96

9 Subramani A and Floudas CA. (2012) Structure Prediction of Loops with Fixed and Flexible Stems. J Phys Chem B **116**, 6670-6682

10 Rayan A. (2010) New vistas in GPCR 3D structure prediction. J Mol Model **16**, 183-191

11 Sellers BD, Nilmeier JP and Jacobson MP. (2010) Antibodies as a model system for comparative model refinement. Proteins **78**, 2490-2505

12 Rayan A, Senderowitz H and Goldblum A. (2004) Exploring the conformational space of cyclic peptides by a stochastic search method. J Mol Graph Model **22**, 319-333

13 Rayan A, Noy E, Chema D et al. (2004) Stochastic algorithm for kinase homology model construction. Curr Med Chem **11**, 675-692

14 Burke DF, Deane CM and Blundell TL. (2000) Browsing the SLoop database of structurally classified loops connecting elements of protein secondary structure. Bioinformatics **16**, 513-519

15 Coutsias EA, Seok C, Jacobson MP et al. (2004) A kinematic view of loop closure. J Comput Chem **25**, 510-528

16 DePristo MA, de Bakker PI, Lovell SC et al. (2003) Ab initio construction of polypeptide fragments: efficient generation of accurate, representative ensembles. Proteins **51**, 41-55

17 Canutescu AA and Dunbrack RL., Jr. (2003) Cyclic coordinate descent: A robotics algorithm for protein loop closure. Protein Sci **12**, 963-972

18 Go M, Go N and Scheraga HA. (1970) Molecular theory of the helix-coli transition in polyamino acids. II. Numerical evaluation of s and sigma for polyglycine and poly-L-alaine in the absence (for s and sigma) and presence (for sigma) of solvent. J Chem Phys **52**, 2060-2079

19 Moult J and James MN. (1986) An algorithm for determining the conformation of polypeptide segments in proteins by systematic search. Proteins **1**, 146-163

20 Shenkin PS, Yarmush DL, Fine R et al. (1987) Predicting antibody hypervariable loop conformation. I. Ensembles of random conformations for ringlike structures. Biopolymers **26**, 2053-2085

21 Xiang Z, Soto CS and Honig B. (2002) Evaluating conformational free energies: the colony energy and its application to the problem of loop prediction. Proc Natl Acad Sci USA **99**, 7432-7437

22 Sudarsanam S, DuBose RF, March CJ et al. (1995) Modeling protein loops using a phi i + 1, psi i dimer database. Protein Sci **4**, 1412-1420

23 Jacobson MP, Pincus DL, Rapp CS et al. (2004) A hierarchical approach to all-atom protein loop prediction. Proteins **55**, 351-367

24 Engh RA and Huber RCA. (1991) Structure Prediction of Loops with Fixed and Flexible Stems. Acta Cryst. **A47**, 392–400

25 Touw WG and Vriend G. (2010) On the complexity of Engh and Huber refinement restraints: the angle tau as example. Acta Crystallogr D Biol Crystallogr **66**, 1341-1350

26 Morris AL, MacArthur MW, Hutchinson EG et al. (1992) Stereochemical quality of protein structure coordinates. Proteins **12**, 345-364

27 Weiner SJ, Kollman PA, Case DA et al. (1984) A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. J. Am. Chem. Soc. **106** 765-784

© 2012; AIZEON Publishers

This is an Open Access article distributed under the terms of the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.