BIOINFORMATICS ORIGINAL PAPER

Vol. 27 no. 6 2011, pages 785–790
doi:10.1093/bioinformatics/btr009

*Structural bioinformatics*

Advance Access publication January 6, 2010

# Improved side-chain modeling by coupling clash-detection guided iterative search with rotamer relaxation

Yang Cao[1,2,†], Lin Song[1,2,†], Zhichao Miao[1,2], Yun Hu[1,2], Liqing Tian[1,2] and Taijiao Jiang[1,*]

[1]National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101 and [2]Graduate University of Chinese Academy of Sciences, Beijing 100039, China

Associate Editor: Martin Bishop

## ABSTRACT

**Motivation:** Side-chain modeling has seen wide applications in computational structure biology. Most of the popular side-chain modeling programs explore the conformation space using discrete rigid rotamers for speed and efficiency. However, in the tightly packed environments of protein interiors, these methods will inherently lead to atomic clashes and hinder the prediction accuracy.

**Results:** We present a side-chain modeling method (CIS-RR), which couples a novel clash-detection guided iterative search (CIS) algorithm with continuous torsion space optimization of rotamers (RR). Benchmark testing shows that compared with the existing popular side-chain modeling methods, CIS-RR removes atomic clashes much more effectively and achieves comparable or even better prediction accuracy while having comparable computational cost. We believe that CIS-RR could be a useful method for accurate side-chain modeling.

**Availability:** CIS-RR is available to non-commercial users at our website: http://jianglab.ibp.ac.cn/lims/cisrr/cisrr.html.

**Contact:** taijiao@moon.ibp.ac.cn

**Supplementary information:** Supplementary data are available at *Bioinformatics* online.

Received on October 23, 2010; revised on December 30, 2010; accepted on January 3, 2011

## 1 INTRODUCTION

Side-chain modeling is the prediction of a protein's side-chain conformation for a given backbone structure. Due to its importance in protein design (Dahiyat and Mayo, 1997), point mutation analysis (Feyfant *et al*., 2007), protein–protein docking (Wang *et al*., 2005), homology modeling (Fiser *et al*., 2002) and other aspects of structural modeling, side-chain modeling has been intensively studied for many years. A thorough search through the continuous side-chain conformations is inefficient and not feasible given the existing search algorithms and current computational power. Luckily, side-chain dihedral angles are not evenly distributed, which occur in tight clusters around certain values (Benedetti *et al*., 1983; Bhat *et al*., 1979; Chandras and Ramachan, 1970; Janin *et al*., 1978; Ponder and Richards, 1987). Based on these clusters, the side-chain conformational space can be discretized into rotamers which represent the conformations of these clusters, and an ensemble of rotamers for all side-chain types is thus called a rotamer library (Dunbrack, 2002; Dunbrack and Cohen, 1997; Dunbrack and Karplus, 1993, 1994). Based on rotamer definition, side-chain modeling can be efficiently transformed into a combinatorial search problem which is to find the best rotamer combination from all possible combinations of side-chain rotamers. For this methodology, many efforts have attempted including development of powerful scoring functions (Eyal *et al*., 2004; Jacobson *et al*., 2002; Liang and Grishin, 2002; Lu *et al*., 2008; Peterson *et al*., 2004), well-sampled rotamer libraries (DeMaeyer *et al*., 1997; Dunbrack and Cohen, 1997; Krivov *et al*., 2009; Xiang and Honig, 2001) and effective search strategies (Canutescu *et al*., 2003; Desmet *et al*., 1992, 2002; Hwang and Liao, 1995; Kingsford *et al*., 2005; Koehl and Delarue, 1995; Leach and Lemon, 1998; Liu *et al*., 2003; Xu, 2005). So far, rotamer-based side-chain modeling methods have become the most popular ones. However, because the side-chain conformations are approximated by discrete rotamers, these methods could inherently cause atomic clashes in the modeled structures, preventing them from accurate modeling of protein side chains. Furthermore, atomic clashes can lead to high energy and instability of structures, and thus the elimination of atomic clashes is beneficial for the subsequent analysis such as protein–protein docking (Wang *et al*., 2005; Zhang and Duan, 2006). Atomic clashes can be alleviated to some extent by utilizing a very large rotamer library (Peterson *et al*., 2004; Xiang and Honig, 2001) or expanding a rotamer into finely sampled subrotamers (Krivov *et al*., 2009; Mendes *et al*., 1999). Other more sophisticated approaches could include an additional step of rotamer refinement by off-rotamer space sampling (Krieger *et al*., 2009; Wang *et al*., 2005).

In our study, to achieve a more effective side-chain modeling, we attempt to integrate the advances from side-chain modeling in both rotamer-based search and rotamer optimization. Our method, called CIS-RR, performs clash-detection guided iterative search (CIS) of side-chain rotamers while continuously optimizes side-chain conformations using a conjugate gradients method. Careful implementation of CIS-RR shows obvious improvements of prediction accuracy and elimination of atomic clashes, in contrast with the traditional iterative search method. Further benchmark testing shows, compared with the existing popular side-chain modeling programs, CIS-RR removes atomic clashes much more effectively and achieves comparable or even better prediction

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

accuracy while having comparable computational cost. The methods and results will be elaborated in the following sections.

## 2 METHODS

### 2.1 Rotamer library

In this work, we use Dunbrack backbone-dependent rotamer library (version of May 2002) (Dunbrack and Cohen, 1997; Dunbrack and Karplus, 1993). We also obtained the protein's standard bond lengths and angles from CHARMM (Brooks *et al*., 1983). Rotamers with a probability $<1\%$ are ignored.

### 2.2 Scoring function

The scoring function was adapted from that of SCWRL3 (Canutescu *et al*., 2003), which consists of two terms:

$$E = E_{\text{vdW}} + k_{\text{rot}} E_{\text{rot}} \tag{1}$$

$E_{\text{vdW}}$ is an empirical van der Waals potential. SCWRL3 only includes a piecewise linear repulsive term. In this work, a quadratic attractive term is also included. For a pair of atoms,

$$E_{\text{vdw}}(i,j) = \begin{cases} k_{\text{rep}}(1 - r_{ij}/R_{ij}) & \text{for } r_{ij} \le R_{ij} \\ k_{\text{att}}\left[(r_{ij}/R_{ij})^2 - 3.0(r_{ij}/R_{ij}) + 2.0\right] & \text{for } R_{ij} < r_{ij} < 2.0R_{ij} \\ 0.0 & \text{for } r_{ij} \ge 2.0R_{ij} \end{cases} \tag{2}$$

where $r_{ij}$ is the interatomic distance between atoms $i$ and $j$ and $R_{ij}$ is the sum of their effective van der Waals radii, which are obtained from Chothia's work (Chothia, 1976) with a 5% reduction. The hydrogen atoms are ignored.

$E_{\text{rot}}$ is the rotamer term, which measures the preferences of the side-chain conformers. For a rotamer $r_m$ of residue $m$,

$$E_{\text{rot}}(r_m) = \gamma\left(-\log\left(\frac{p(r_m|\phi,\varphi)}{p(\max|\phi,\varphi)}\right) + k_{\text{tor}}\sum_n\left(\frac{\chi_n(r_m) - \chi_{lib\_n}}{\sigma_{lib\_n}}\right)^2\right) \tag{3}$$

where $p(r_m|\phi,\varphi)$ is the rotamer library probability for the given backbone dihedral angle $\phi$ and $\varphi$. It is normalized to the maximum probability $p(\max|\phi,\varphi)$ of residue $m$ in the same backbone dihedral angles (Canutescu *et al*., 2003). We also considered a second term to penalize the drifting of side-chain dihedral angles away from the nearest rotamer library values. The side-chain dihedral angle $\chi_n(r_m)$ is the $n$-th $\chi$ dihedral angle of $r_m$ $\chi_{lib\_n}$ and $\sigma_{lib\_n}$ are the corresponding rotamer library dihedral angle and its standard deviation. Since aromatic side-chains are subject to have clashes, their $E_{\text{rot}}$ could be underestimated compared with their $E_{\text{vdW}}$. Thus the scaling factor $\gamma$ is introduced and set as follows:

$$\gamma = \begin{cases} 2, & \text{if } r_m \in \{\text{His, Phe, Tyr, Trp}\} \\ 0, & \text{if } r_m \in \{\text{Gly, Ala}\} \\ 1, & \text{otherwise} \end{cases} \tag{4}$$

The parameters ($k_{\text{rot}} = 0.35$, $k_{\text{rep}} = 5.882$, $k_{\text{att}} = 0.08$ and $k_{\text{tor}} = 0.4$) of the scoring function were obtained by training on 55 high-resolution crystal structures that were randomly selected from the Protein Data Bank (PDB; Berman *et al*., 2000). We followed the training procedure used by former works (Liang and Grishin, 2002; Petrella *et al*., 1998; Wilson *et al*., 1993; Xiang *et al*., 2007). In brief, parameters were systematically optimized by searching the best conformation of one residue each time while keeping the others fixed in their native conformations.

### 2.3 Disulfide bonds

Before performing the scoring function calculation, CIS-RR uses the formula $E_{ss} = 0.1 \cdot S - 4.0$ to evaluate the cysteine pair, which forms a disulfide bond. $S$ is the disulfide bond score as used in SCWRL3 (Canutescu *et al*., 2003). If the side chains of the cysteine pairs form disulfide bonds ($E_{ss} < 0$), $E_{ss}$ will be added to the scoring function.

```
Generate initial conformation using the rotamers with the highest library probabilities
en = energy of initial conformation
For each residue i
    For each rotamer r of i
        Run conjugate gradients minimization for r
        If r clashes with other residues
            For each clashed residue j
                For each rotamer s of j
                    Run conjugate gradients minimization for s and r
                    If new en < en
                        en = new en
                        Store r and s
                End-for s
            End-for j
        Else
            If new en < en
                en = new en
                Store r
    End-for r
    Update conformation of r to be i or (r, s) to be (i, j)
End-for i
```

**Fig. 1.** Pseudocodes of CIS-RR algorithm. The complexity of the algorithm is $O(np^2)$, where $n$ is the number of residues; $p$ is the number of rotamers per residue.

### 2.4 Rotamer relaxation

In this work, rotamer relaxation (RR) is implemented by using a conjugate gradients method to optimize the side-chain dihedral angles near the rotamer value. The conjugate gradients method follows Fletcher–Reeves algorithm (Fletcher and Reeves, 1964). To produce a set of directions to rotate a side-chain, the directions are calculated from the derivatives of the scoring function with respect to side-chain dihedral angles:

$$E' = E'_{\text{vdW}} + k_{\text{rot}} E'_{\text{rot}} \tag{5}$$

The derivative of the van der Waals potential for each of the dihedral angle $\chi_n$ is

$$E'_{\text{vdW}}(\chi_n) = \sum_i\sum_j \frac{\partial E_{\text{vdW}}}{\partial r_{ij}}\left(\frac{\partial r_{ij}}{\partial x_i}\vec{i} + \frac{\partial r_{ij}}{\partial y_i}\vec{j} + \frac{\partial r_{ij}}{\partial z_i}\vec{k}\right) \bullet \frac{\vec{s}_{\text{axis}} \times \vec{s}_i}{\left\|\vec{s}_{\text{axis}} \times \vec{s}_i\right\|}\left\|\vec{s}_i\right\| \tag{6}$$

where $x_i, y_i, z_i$ is the Cartesian coordinates of atom $i$ constrained by $\chi_n$. $\vec{s}_{\text{axis}}$ is the unit vector of the torsion axis of $\chi_n$. $\vec{s}_i$ is the perpendicular vector from atom $i$ to the torsion axis $\vec{s}_{\text{axis}}$ (Supplementary Fig. S1 and Table S1). The detailed derivation of the above equation is provided in Supplementary Material.

The derivative of the rotamer term for each of the dihedral angle $\chi_n$ is

$$E'_{\text{rot}}(\chi_n) = 2k_{\text{tor}}\gamma\frac{\chi_n(r_m) - \chi_{lib\_n}}{\sigma_{lib\_n}} \tag{7}$$

### 2.5 CIS-RR

The CIS-RR operates as follows:

First, the starting side-chain conformation of each residue is constructed by the rotamers with the highest probability at each position. Then, for each residue $i$, every one of its rotamers will be optimized as described in the Section 2.4 and tested for clashes with other residues that are kept fixed. If a rotamer $r$ has clashes with other residues (i.e. the score of the van der Waals term $>1.0$), rotamer $r$ is fixed temporarily and the clashing residues will change to each of the other rotamers, proceeding with RR. After all rotamers of the residue $i$ are explored, the side-chain conformations of residue $i$ or residue $i$ and its clashing residues will be updated with side-chain dihedral angles that generate the lowest score during the search. The more detailed process is described as pseudocodes (Fig. 1). The above process keeps iterating from the most exposed residue to the least until the score converges to a stable value.

Clash-detection step plays an important role in CIS-RR. According to our analysis, ∼46% tested rotamers have clashes with other residues. Detailed information is provided in Supplementary Table S2.

For CIS without RR, the RR process is bypassed in the above procedure.

### 2.6 Traditional iterative search coupled with RR

Traditional iterative search coupled with RR (TIS-RR) tests every rotamer for a given residue while keeping other residues fixed. All the rotamers are subjected to RR and the one resulting in the lowest score will replace the original one. It updates one residue at a time and keeps looping until the side chains stop changing conformations.

For TIS without RR, the RR process is bypassed in the above procedure.

### 2.7 Evaluation methods

The modeling accuracy is defined as the percentage of correctly predicted side-chain $\chi_1$ and $\chi_{1+2}$ dihedral angles within a threshold of 40° compared with the crystal structures (Bower *et al*., 1997; Eyal *et al.*, 2004; Liang and Grishin, 2002). In our work, the modeling accuracy is also indicated by root mean square deviation (RMSD) of side-chain heavy atoms. It calculates the square root of the mean of the square distances over all atoms from the predicted structures to the crystal ones. Before evaluating accuracy, the crystallographically symmetric terminal groups of Asp, Glu, Phe, Arg and Tyr are flipped to yield optimal atom matching. For Asn, Gln and His, due to the difficulty to distinguish different terminal atoms, their terminal groups are also flipped (Eyal *et al*., 2004).

The atomic clash is also used to evaluate the modeling structures. If two atoms over two bonds away have the distance <60% of the sum of their respective van der Waals radii, they are regarded as an atomic clash.

### 2.8 Benchmark testing

Two different test sets have been used in our study. The first test set consists of 180 crystal structures, which were obtained from the test set of SCWRL3 (Canutescu *et al*., 2003). It was used to examine the contributions of the two components of CIS-RR, namely CIS and RR, to the side-chain modeling. For a fair comparison, CIS-RR, CIS, TIS-RR and TIS all start from the same initial side-chain conformations in the testing. The second test set was used to compare CIS-RR with several popular side-chain modeling programs. It consists of 65 high-resolution crystal structures that were used in previous side-chain modeling studies (Harder *et al*., 2010; Jain *et al*., 2006; Lu *et al*., 2008; Peterson *et al*., 2004). All the tests were run on an Intel Q9550 processor.

### 2.9 Software availability

CIS-RR is implemented in C++ and has been complied on Linux system. The executable version of our method is now available to non-commercial users at http://jianglab.ibp.ac.cn/lims/cisrr/cisrr.html.

## 3 RESULTS

In this section, we will first give an overview of our method CIS-RR. Then we will analyze the contributions of CIS and RR to the improvement of side-chain modeling and further investigate how CIS-RR effectively eliminates clashes. Finally, to gain an overall performance of CIS-RR, we compare it to several existing popular side-chain modeling programs.

### 3.1 Overview of CIS-RR

CIS-RR is a side-chain modeling program that performs CIS and RR simultaneously (Section 2 and Fig. 1). Previously, TIS based on single residues has been attempted in side-chain modeling, which each time simply searches all rotamers of a given residue for the best one while keeping other residues fixed (Dunbrack and Karplus, 1993; Xiang and Honig, 2001). This method is relatively fast and able to couple with RR easily. However, it is sensitive to the former states and tends to get trapped in local minima, which hinders the modeling accuracy. To efficiently improve this method, we employ a clash-detection step to help side chains jump out of the local minima that is blocked by the mispacked surrounding residues. The implementation of CIS can not only effectively improve accuracy and resolve clashes but also achieve a desirable speed in side-chain modeling, for which we will show the results later.

RR samples off-rotamer side-chain conformations quite effectively, but it usually cannot produce large enough changes of the dihedral angles to correct the mispacked rotamers. In the pioneer work from Baker lab, Wang *et al*. implemented a method, called RTMIN, which couples TIS with RR (i.e. side-chain minimization) to optimize side chains (Krieger *et al*., 2009; Wang *et al.*, 2005). Inspired by their work, we couple the improved search algorithm CIS with RR and find that CIS-RR is superior to TIS-RR for its effectively removing clashes and significantly improving the accuracy of side-chain modeling.

### 3.2 The CIS and RR both contribute to the improvement of side-chain modeling over TIS

First, we compare CIS with TIS in side-chain modeling. For ease of comparison, both methods searched through the same rotamer library without RR. As shown in Table 1, compared with TIS, CIS dramatically reduces the number of atomic clashes from 224 to 125 out of 180 proteins while incurring a moderate increase in time consumption. Moreover, the accuracy of side-chain modeling has also been improved by CIS: the $\chi_1$ and $\chi_{1+2}$ of the whole proteins have been improved 1.3% (84.7% by CIS versus 83.4% by TIS) and 1.7% (73.8% by CIS versus 72.1% by TIS), respectively, and the RMSD has decreased 0.1 Å (from 1.78 Å to 1.68 Å).

Then we investigate to what extent RR contributes to side-chain modeling by comparing TIS-RR to TIS. Table 1 shows that RR effectively reduces the number of clashed residues from 224 to 63 out of 180 proteins. Moreover, the $\chi_1$ and $\chi_{1+2}$ have also been improved 0.5% (83.9% TIS-RR versus 83.4% TIS) and 1.5% (73.6% TIS-RR versus 72.1% TIS), respectively, for all residues; and the RMSD has decreased 0.08 Å (from 1.78 Å to 1.70 Å). However, RR is a time-consuming step, which incurs a significant increase of computation time (3–5 times of that without RR).

The above analyses clearly show that both CIS and RR can improve side-chain modeling not only in the elimination of atomic clashes but also in the accuracy of side-chain modeling in terms of number of correct $\chi$ dihedral angles and RMSD. Remarkably, when both CIS and RR are integrated into the framework of CIS-RR, the side-chain modeling can be further improved. For the elimination of residue clashes, CIS-RR leaves only 15 clashed residues unresolved in the 180 proteins tested, which is much better than the use of RR or CIS alone (Table 1). For the accuracy improvement in terms of correct $\chi$ dihedral angles and RMSD, CIS-RR causes the $\chi_1$ and $\chi_{1+2}$ improvement of 2.3% and 3.8% (from 83.4% to 85.7% and 72.1% to 75.9%), RMSD decrease of 0.21 Å (from 1.78 Å to 1.57 Å), greater than the improvement summation of CIS (1.3%, 1.7%, 0.1 Å) and TIS-RR (0.5%, 1.5%, 0.08 Å). Taking D-ribose-binding protein (PDB ID: 2DRI) as an example, the percent correct for $\chi_1$ and $\chi_{1+2}$

**Table 1.** Comparison of CIS-RR, TIS-RR, CIS and TIS in the 180-crystal structure set

| Method | Time (min) | Buried | | | All | | | No. of clashes |
|---|---|---|---|---|---|---|---|---|
| | | $\chi_1$ (%) | $\chi_{1+2}$ (%) | RMSD $\pm \sigma$ (Å) | $\chi_1$ (%) | $\chi_{1+2}$ (%) | RMSD $\pm \sigma$ (Å) | |
| TIS | 7 | 89.7 | 81.1 | $1.22 \pm 0.50$ | 83.4 | 72.1 | $1.78 \pm 0.33$ | 224 |
| CIS | 11 | 91.5 | 83.5 | $1.07 \pm 0.41$ | 84.7 | 73.8 | $1.68 \pm 0.32$ | 125 |
| TIS-RR | 28 | 90.0 | 82.6 | $1.12 \pm 0.48$ | 83.9 | 73.6 | $1.70 \pm 0.35$ | 63 |
| CIS-RR | 54 | 92.3 | 85.9 | $0.92 \pm 0.34$ | 85.7 | 75.9 | $1.57 \pm 0.34$ | 15 |

Residues are considered to be buried when there is a <20% solvent-accessibility ratio.

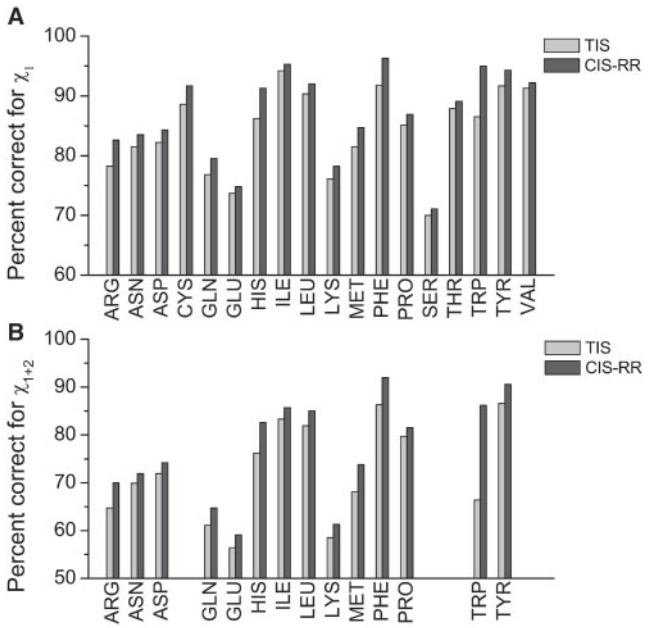


**Fig. 2.** Improvements of CIS-RR over TIS on different amino acids. (**A**) Percent correct within 40° for the $\chi_1$ angle. (**B**) Percent correct within 40° for both the $\chi_1$ and $\chi_2$ angles. The test data used the 180-crystal structure set. Light and dark gray columns indicate the test results of TIS and CIS-RR, respectively.

is 89.3% and 78.6% by using CIS-RR, which have been significantly improved, compared with 84.8%, 70.6% by TIS-RR and 87.6%, 75.4% by CIS. This suggests that CIS and RR, when integrated into CIS-RR, collectively improve the performance of side-chain modeling.

### 3.3 Analysis of the performance of CIS-RR on individual residues

Next, we sought to find out how CIS-RR works on different types of amino acids. Figure 2 shows that CIS-RR improves the percent correct of both $\chi_1$ and $\chi_{1+2}$ dihedral angles of all amino acids. The improvements are more significant for the amino acids with aromatic side-chains (Phe His, Trp and Tyr). For example, the $\chi_{1+2}$ improvement of Trp is up to almost 20% (Fig. 2 and Supplementary Table S3). Our further analysis showed that for side-chain modeling by TIS, the four residues constitute about half of the clashes due to their large-size terminal groups (Supplementary Table S3). In CIS-RR the clashes of these four residues can be significantly reduced to seven clashes (Supplementary Table S3). Therefore, by focusing on elimination of clashes, our method CIS-RR efficiently corrects the mispacked conformations and thus improves the side-chain modeling, particularly for those of large size that is easy to cause clashes in the well-packed proteins.

### 3.4 Comparison of CIS-RR with the existing popular side-chain modeling programs

To gain an overall performance of CIS-RR, we further tested it on another popular test dataset consisting of 65 high-resolution crystal structures by making comparisons with several popular side-chain modeling programs, SCWRL4 (Krivov *et al*., 2009), IRECS (Hartmann *et al*., 2007), Sccomp (Eyal *et al*., 2004) and SCAP (Xiang and Honig, 2001). SCWRL is probably the most widely used program in the field. The latest version is SCWRL4, which is improved from SCWRL3 (Canutescu *et al*., 2003) by using the new rotamer library, more efficient search algorithms and a soft van der Waals potential plus hydrogen bonding based scoring function. IRECS does an efficient ensemble selection of most probable side-chain conformations using a knowledge-based statistical potential. Sccomp takes into account solvent accessibility and contact surfaces in the scoring function for accurate modeling. SCAP employs the iterative procedure and CHARMM force field to search through a very large rotamer library. All those methods are based on discrete rotamers.

Table 2 shows the test results of the five programs, in which CIS-RR achieved prediction accuracies of 86.4% and 76.7% for $\chi_1$ and $\chi_{1+2}$, respectively (see Supplementary Table S4 for the predicted accuracy of each protein). These results illustrate that, in terms of percent correct for dihedral angles, CIS-RR is comparable with SCWRL4, and better than IRECS, Sccomp and SCAP, especially for $\chi_{1+2}$ (~5% higher accuracy). Notably, in terms of number of clashes, CIS-RR eliminates almost all the clashes (2 clashes in 65 proteins) and outperforms the other four programs. However, due to the time consuming RR process, the run time of CIS-RR (15 min for 65 proteins) is longer than SCWRL4, IRECS and Sccomp.

We further compared the performance of CIS-RR and SCWRL4 over individual amino acids. As shown in Figure 3 and Supplementary Table S5, among 18 types of amino acids (excluding Ala and Gly), CIS-RR showed better performance for 13 amino acids, especially for Cys, Trp and Met. This can be attributed to the efficient sampling of rotamer and off-rotamer conformations by CIS-RR. However, for the short polar amino acids (Asp, Asn and Ser), CIS-RR shows lower performance, which could be due

**Table 2.** Comparison of CIS-RR with several popular side-chain modeling programs on a 65-crystal structure set

| Method | Time (min) | Buried | | | All | | | No. of clashes | Ref |
|---|---|---|---|---|---|---|---|---|---|
| | | $\chi_1$ (%) | $\chi_{1+2}$ (%) | RMSD $\pm\sigma$ (Å) | $\chi_1$ (%) | $\chi_{1+2}$ (%) | RMSD $\pm\sigma$ (Å) | | |
| CIS-RR | 15 | 93.8 | 90.0 | $0.86 \pm 0.28$ | 86.4 | 76.7 | $1.50 \pm 0.26$ | 2 | This work |
| SCWRL4 | 9 | 92.9 | 89.2 | $0.96 \pm 0.33$ | 85.8 | 76.3 | $1.60 \pm 0.25$ | 39 | Krivov *et al*. (2009) |
| IRECS | 12 | 92.1 | 83.6 | $1.06 \pm 0.26$ | 84.3 | 71.8 | $1.65 \pm 0.25$ | 135 | Hartmann *et al*. (2007) |
| Sccomp | 5 | 92.1 | 86.5 | $1.04 \pm 0.37$ | 83.2 | 72.0 | $1.63 \pm 0.28$ | 11 | Eyal *et al.* (2004) |
| SCAP | 52 | 92.6 | 86.7 | $0.98 \pm 0.36$ | 83.3 | 71.3 | $1.58 \pm 0.24$ | 11 | Xiang and Honig (2001) |

All the programs are used under their default settings. Residues are considered to be buried when there is a <20% solvent-accessibility ratio.

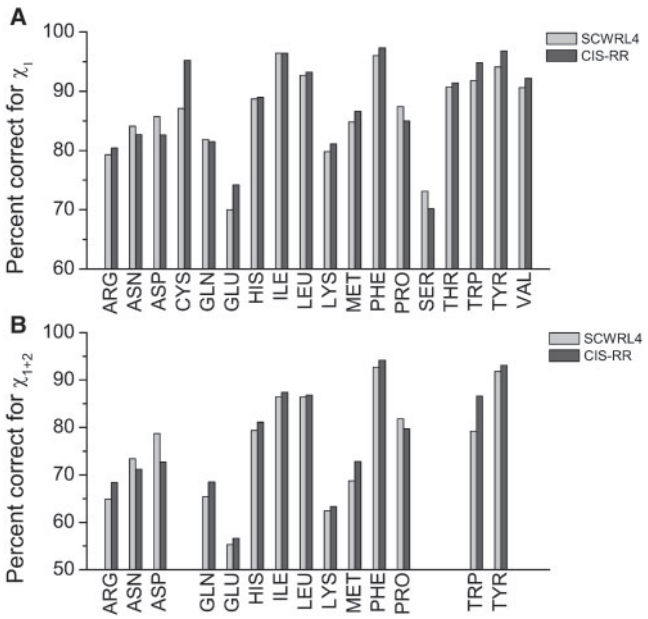


**Fig. 3.** Comparison of CIS-RR and SCWRL4 over different amino acids. (**A**) Percent correct within 40° for the $\chi_1$ angle. (**B**) Percent correct within 40° for both the $\chi_1$ and $\chi_2$ angles. The test data used the 65-crystal structure set. The results for SCWRL4 and CIS-RR are shown in light and dark gray, respectively.

to their difference in scoring functions. Compared with SCWRL4, CIS-RR does not include the hydrogen bonding effects in the scoring function. Supplementary Table S5 also illustrates that CIS-RR eliminates their clashes more effectively, particular for these residues with large-size side chains as described above. Taken together, we argue that CIS-RR and SCWRL4 are complementary approaches by nature in side-chain modeling, which suggests the significance of our work by focusing on effective clash elimination.

## 4 CONCLUSION

In order to improve the discrete rigid rotamer model, we developed a side-chain modeling program that integrates clash-detection guided iterative search with RR in the framework CIS-RR. Our analysis shows that the good performance of CIS-RR in side-chain modeling is mainly due to its effectiveness in clash elimination, which is achieved by extensive sampling of side-chain conformation space through the coupling of CIS and RR. Since CIS-RR is able to effectively relieve atomic clashes in the modeled structures with a desirable speed, it would be a very useful method for accurate side-chain modeling that can complement the current existing methods for subsequent analysis and refinement of protein structures.

## ACKNOWLEDGEMENTS

We would like to thank Dr Aiping Wu and Mr Wentao Dai for their stimulating discussions and Mr Jingzhou Chen and Ms Cuixia Chen for the help in building the web site.

*Funding*: Bai Ren Project of Chinese Academy of Sciences and grants from the Chinese Ministry of Science and Technology; the National Science and Technology Key Project (2008ZX10004-013) and Project 973 (2009CB918503) (to T.J.).

*Conflict of Interest*: none declared.

## REFERENCES

Benedetti,E. *et al.* (1983) Statistical and energetic analysis of side-chain conformations in oligopeptides. *Int. J. Pep.t Prot. Res.*, **22**, 1–15.

Berman,H. *et al.* (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.

Bhat,T.N. *et al.* (1979) Analysis of side-chain conformation in proteins. *Int. J. Pept. Prot. Res.*, **13**, 170–184.

Bower,M.J. *et al.* (1997) Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: A new homology modeling tool. *J. Mol. Biol.*, **267**, 1268–1282.

Brooks,B. *et al.* (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, **4**, 187–217.

Canutescu,A.A. *et al.* (2003) A graph-theory algorithm for rapid protein side-chain prediction. *Prot. Sci.*, **12**, 2001–2014.

Chandras,R. and Ramachan,G.N. (1970) Studies on conformation of amino acids .11. Analysis of observed side group conformations in proteins. *Int. J. Prot. Res.*, **2**, 223–233.

Chothia,C. (1976) Nature of accessible and buried surfaces in proteins. *J. Mol. Biol.*, **105**, 1–14.

Dahiyat,B.I. and Mayo,S.L. (1997) Probing the role of packing specificity in protein design. *Proc. Natl Acad. Sci. USA*, **94**, 10172–10177.

DeMaeyer,M. *et al.* (1997) All in one: a highly detailed rotamer library improves both accuracy and speed in the modelling of sidechains by dead-end elimination. *Fold. Des.*, **2**, 53–66.

Desmet,J. *et al.* (1992) The dead-end elimination theorem and its use in protein side-chain positioning. *Nature*, **356**, 539–542.

Desmet,J. *et al.* (2002) Fast and accurate side-chain topology and energy refinement (FASTER) as a new method for protein structure optimization. *Prot. Struct. Funct. Genet.*, **48**, 31–43.

Dunbrack,R.L. (2002) Rotamer libraries in the 21(st) century. *Curr. Opin. Struct. Biol.*, **12**, 431–440.

Dunbrack,R.L. and Cohen,F.E. (1997) Bayesian statistical analysis of protein side-chain rotamer preferences. *Prot. Sci.*, **6**, 1661–1681.

Dunbrack,R.L. and Karplus,M. (1993) Backbone-dependent rotamer library for proteins - application to side-chain prediction. *J. Mol. Biol.*, **230**, 543–574.

Dunbrack,R.L. and Karplus,M. (1994) Conformational-analysis of the backbone-dependent rotamer preferences of protein side-chains. *Nat. Struct. Biol.*, **1**, 334–340.

Eyal,E. *et al.* (2004) Importance of solvent accessibility and contact surfaces in modeling side-chain conformations in proteins. *J. Comput. Chem.*, **25**, 712–724.

Feyfant,E. *et al.* (2007) Modeling mutations in protein structures. *Prot. Sci.*, **16**, 2030–2041.

Fiser,A. *et al.* (2002) Evolution and physics in comparative protein structure modeling. *Acc. Chem. Res.*, **35**, 413–421.

Fletcher,R. and Reeves,C.M. (1964) Function minimization by conjugate gradients. *Comput J.*, **7**, 149–154.

Harder,T. *et al.* (2010) Beyond rotamers: a generative, probabilistic model of side chains in proteins. *BMC Bioinformatics*, **11**, 306.

Hartmann,C. *et al.* (2007) IRECS: a new algorithm for the selection of most probable ensembles of side-chain conformations in protein models. *Prot. Sci.*, **16**, 1294–1307.

Hwang,J.K. and Liao,W.F. (1995) Side-chain prediction by neural networks and simulated annealing optimization. *Prot. Eng.*, **8**, 363–370.

Jacobson,M.P. *et al.* (2002) Force field validation using protein side chain prediction. *J. Phys. Chem. B*, **106**, 11673–11680.

Jain,T.*et al.* (2006) Configurational-bias sampling technique for predicting side-chain conformations in proteins. *Prot. Sci.*, **15**, 2029–2039.

Janin,J. *et al.* (1978) Conformation of amino-acid side-chains in proteins. *J. Mol. Biol.*, **125**, 357–386.

Kingsford,C.L. *et al.* (2005) Solving and analyzing side-chain positioning problems using linear and integer programming. *Bioinformatics*, **21**, 1028–1036.

Koehl,P. and Delarue,M. (1995) A self-consistent mean-field approach to simultaneous gap closure and side-chain positioning in homology modeling. *Nat. Struct. Biol.*, **2**, 163–170.

Krivov,G.G. *et al.* (2009) Improved prediction of protein side-chain conformations with SCWRL4. *Prot. Struct. Funct. Bioinform.*, **77**, 778–795.

Krieger,E. *et al.* (2009) Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in CASP8. *Prot. Struct. Funct. Bioinform.*, **77**, 114–122.

Leach,A.R. and Lemon,A.P. (1998) Exploring the conformational space of protein side chains using dead-end elimination and the A* algorithm. *Prot. Struct. Funct. Genet.*, **33**, 227–239.

Liang,S.D. and Grishin,N.V. (2002) Side-chain modeling with an optimized scoring function. *Prot. Sci.*, **11**, 322–331.

Liu,Z.J. *et al.* (2003) Beyond the rotamer library: genetic algorithm combined with the disturbing mutation process for upbuilding protein side-chains. *Prot. Struct. Funct. Genet.*, **50**, 49–62.

Lu,M.Y. *et al.* (2008) OPUS-PSP: an orientation-dependent statistical all-atom potential derived from side-chain packing. *J. Mol. Biol.*, **376**, 288–301.

Lu,M.Y. *et al.* (2008) OPUS-Rota: a fast and accurate method for side-chain modeling. *Prot. Sci.*, **17**, 1576–1585.

Mendes,J. *et al.* (1999) Improved modeling of side-chains in proteins with rotamer-based methods: a flexible rotamer model. *Prot. Struct. Funct. Genet.*, **37**, 530–543.

Peterson,R.W. *et al.* (2004) Improved side-chain prediction accuracy using an ab initio potential energy function and a very large rotamer library. *Prot. Sci.*, **13**, 735–751.

Petrella,R.J. *et al.* (1998) Protein sidechain conformer prediction: a test of the energy function. *Fold. Des.*, **3**, 353–377.

Ponder,J.W. and Richards,F.M. (1987) Tertiary templates for proteins - use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol.*, **193**, 775–791.

Wang,C. *et al.* (2005) Improved side-chain modeling for protein-protein docking. *Prot. Sci.*, **14**, 1328–1339.

Wilson,C. *et al.* (1993) Modeling side-chain conformation for homologous proteins using an energy-based rotamer search. *J. Mol. Biol.*, **229**, 996–1006.

Xiang,Z.X. and Honig,B. (2001) Extending the accuracy limits of prediction for side-chain conformations. *J. Mol. Biol.*, **311**, 421–430.

Xiang,Z.X. *et al.* (2007) Prediction of side-chain conformations on protein surfaces. *Prot. Struct. Funct. Bioinform.*, **66**, 814–823.

Xu,J.B. (2005) Rapid protein side-chain packing via tree decomposition. *Res. Comput. Mol. Biol., Lect Notes Comput Sci*, **3500**, 423–439.

Zhang,W. and Duan,Y. (2006) Grow to Fit Molecular Dynamics (G2FMD): an ab initio method for protein side-chain assignment and refinement. *Prot. Eng. Des. Select.*, **19**, 55–65.