PROTEINS: Structure, Function, and Genetics 37:530–543 (1999)

Improved Modeling of Side-Chains in Proteins With Rotamer-Based Methods: A Flexible Rotamer Model

Joaquim Mendes, António M. Baptista, Maria Arménia Carrondo, and Cláudio M. Soares*
Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Oeiras, Portugal

ABSTRACT **Side-chain modeling has a widespread application in many current methods for protein tertiary structure determination, prediction, and design. Of the existing side-chain modeling methods, rotamer-based methods are the fastest and most efficient. Classically, a rotamer is conceived as a single, rigid conformation of an amino acid side-chain. Here, we present a flexible rotamer model in which a rotamer is a continuous ensemble of conformations that cluster around the classic rigid rotamer. We have developed a thermodynamically based method for calculating effective energies for the flexible rotamer. These energies have a one-to-one correspondence with the potential energies of the rigid rotamer. Therefore, the flexible rotamer model is completely general and may be used with any rotamer-based method in substitution of the rigid rotamer model. We have compared the performance of the flexible and rigid rotamer models with one side-chain modeling method in particular (the self-consistent mean field theory method) on a set of 20 high quality crystallographic protein structures. For the flexible rotamer model, we obtained average predictions of 85.8% for χ_1, 76.5% for χ_{1+2} and 1.34 Å for root-mean-square deviation (RMSD); the corresponding values for core residues were 93.0%, 87.7% and 0.70 Å, respectively. These values represent improvements of 7.3% for χ_1, 8.1% for χ_{1+2} and 0.23 Å for RMSD over the predictions obtained with the rigid rotamer model under otherwise identical conditions; the corresponding improvements for core residues were 6.9%, 10.5% and 0.43 Å, respectively. We found that the predictions obtained with the flexible rotamer model were also significantly better than those obtained for the same set of proteins with another state-of-the-art side-chain placement method in the literature, especially for core residues. The flexible rotamer model represents a considerable improvement over the classic rigid rotamer model. It can, therefore, be used with considerable advantage in all rotamer-based methods commonly applied to protein tertiary structure determination, prediction, and design and also in predictions of free energies in mutational studies. Proteins 1999;37:530–543.**

Key words: side-chain modeling; flexible rotamer model; Kirkwood superposition approximation; effective rotamer energy; protein design

INTRODUCTION

Modeling the side-chain conformations of amino acid residues in protein molecules from the residue sequence and a model of the protein backbone is a natural subproblem of many current efforts in protein tertiary structure determination, prediction, and design. On the experimental side, the last stage of protein tertiary structure determination by X-ray crystallography prior to refinement, involves placing side-chains onto a model of the backbone, previously traced through the continuous electron density. On the theoretical side, homology-based modeling (see Martin et al.[1] and references therein), threading algorithms (see Marchler-Bauer and Byrant[2] and references therein), and ab initio methods (see Zemla et al.[3] and references therein) for predicting protein tertiary structure, as well as inverse folding methods for de novo protein design (for example[4–6]) and methods for simulating protein folding (for example[7]), usually provide a model of the protein backbone onto which side-chains must subsequently be placed. In light of this widespread application, modeling of side-chain conformation has been an active area of protein research over the last decade (for a review see Vásquez[8]).

All side-chain modeling methods are based on the definition of a function that measures the proximity of a given conformation of the protein to its true native conformation, and on the optimization of this function by exploration of side-chain conformational space. A key approximation in most methods is the discretization of side-chain conformational space, whereby a side-chain is only allowed to adopt a discrete set of conformations. This approximation is based on the observation that, in high-resolution experimental protein structure models, most side-chains tend to cluster around a discrete set of favored conformations, known as rotamers; in most cases, these were found to correspond to local minima on the side-chain's potential energy map.[9–11] The discretization of side-chain conformational space in these methods leads to a drastic reduction

Grant sponsor: PRAXIS XXI; Grant number: PBIC/C/BIO/2037/95; Grant sponsor: EC; Grant number: BIO04 CT96-0413. Grant sponsor: PRAXIS XXI; Fellowship numbers: BD/2740/94, BPD/4151/94 and BIC/1921/97.
*Correspondence to: Dr. Cláudio M. Soares, Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Av. da República, Apartado 127, 2781-901 Oeiras, Portugal. E-mail: claudio@itqb.unl.pt
Received 9 April 1999; Accepted 6 July 1999

in search space, compared to methods in which side-chain conformational space is explored in a continuous (for example[12]) or quasi-continuous (for example[13,14]) fashion. As a consequence, the speed and efficiency with which discrete space methods find close to optimal protein conformations is dramatically enhanced in relation to continuous space methods.

In spite of the enormous advantages of speed and efficiency that result from the rotamer approximation, it also suffers from two major drawbacks. On one hand, even if the correct rotamer† for a given side-chain exists in the rotamer library, this rotamer may deviate substantially from the true average side-chain conformation. On the other hand, a rotamer represents a rigid conformation of the side-chain whereas, in a real protein, side-chains are flexible, as has been shown by a wide range of techniques.[15,16] Focusing only on energy-based side-chain modeling methods, both these limitations result in errors in the side-chain's energy and, consequently, in errors in side-chain prediction.

In relation to the first drawback, the deviation of the correct rotamer from the true average side-chain conformation means that, even when the side-chain is modeled with the correct rotamer, it will have errors in its atomic coordinates. These errors evidently result in errors in all of the side-chain's potential energy terms: its intrinsic potential energy, its potential energy of interaction with the backbone, and its potential energy of interaction with the side-chains of other residues. Of special importance are the errors in the non-bonded van der Waals potential energy, commonly measured by a Lennard-Jones potential. The error in this potential is particularly severe if there is steric overlap between atoms. Various approaches have been attempted to reduce this type of error. One approach that has met with relative success, and that is now almost universally implemented in energy-based rotamer methods, is the truncation of the Lennard-Jones potential to some fixed value for short interatomic distances (for example[13,17–19]). Although this approach solves the problem of steric overlap to a large degree, it does not work very well when there are many overlapping atoms. Moreover, it does not solve the more general problem of the deviation of the correct rotamer from the true average side-chain conformation, since the coordinates of even non-overlapping atoms are still subject to error. Other attempts to reduce this deviation have been the generation of additional rotamers with torsion angles at about 1 standard deviation (SD) above or below those of the standard rotamer,[18] the replacement of the usual rotamer-rotamer energies by the minimum energy over a set of conformations around each standard rotamer[18] and the construction of position-specific rotamers by minimization of each standard rotamer in the presence of the rigid backbone.[20] The authors report that none of these approaches met with significant success. However, Lazar and co-workers[21] have reported two approaches that have met with success in the context of protein design, which are similar to those described by Holm and Sander[18] referred to above.

In relation to the second drawback, since a rotamer represents a rigid conformation of a side-chain, its potential energy is a static quantity that does not reflect the dynamic nature of the interactions of a side-chain of a real protein. Indeed, since real side-chains are flexible, the potential energy of a side-chain is a dynamic quantity that depends on the side-chain's instantaneous conformation and on the instantaneous conformation of all other side-chains. A correct measure of the interactions of a side-chain would be an effective energy reflecting this dynamic nature by taking into account all instantaneous potential energy values. Importantly, such an effective energy would contain not only energetic terms but also entropic terms.

Here we present a novel method for calculating effective energies that addresses both of the above-mentioned problems inherent to rotamer-based methods. The basis of this method is a flexible rotamer model. In this model, each rotamer is represented not by a single, rigid conformation, but rather by a continuous ensemble of conformations that cluster around the standard rigid rotamer. We show that the effective energy of a particular configuration of flexible rotamers is given by a conditional free energy that correctly reflects the ensemble of conformations associated with each flexible rotamer. Using a standard thermodynamic approximation, this global free energy can be decomposed into effective intrinsic and pairwise energy terms for each rotamer—effective intrinsic energy, effective energy of interaction with the backbone and effective energy of interaction with the rotamers of other residues—that are analogous to the respective potential energy terms into which the total potential energy is commonly decomposed in rotamer-based methods. Therefore, the method we present is completely general, and the effective energy terms it provides can be used in combination with any of the search strategies commonly used in rotamer-based methods to find protein conformations of low-lying energy, such as the self-consistent mean field theory (SCMFT) method,[19] the Monte Carlo algorithm,[18] the dead-end elimination method,[22] among others.

We have used the flexible rotamer model with the SCMFT method for modeling side-chain conformation[19] incorporating the recently introduced multiple run random initialization (MRRI) protocol.[23] However, we stress that this choice of side-chain modeling method is completely arbitrary. We compare the results with those obtained using the rigid rotamer model and with those of the side-chains with a rotamer library (SCWRL) method.[24] We also tested the effect on prediction accuracy of variations in several factors upon which the effective energy calculation depends. All results were in accordance with what would be expected on a physical basis.

METHODS

Theory

The method for calculating effective rotamer energies, that we will now describe, is based on a flexible rotamer

†A rotamer whose torsion angles all belong to the same local potential energy minimum as the corresponding torsion angles of the true average side-chain conformation.

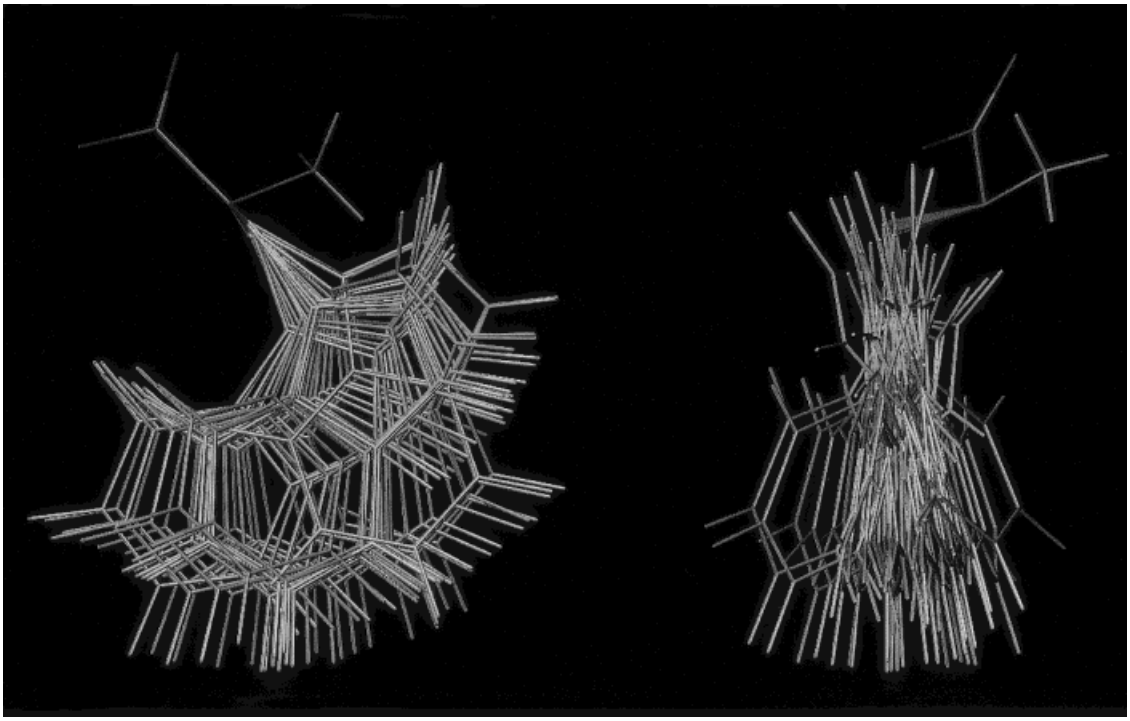

Fig. 1. The flexible rotamer model: two orthogonal views of a flexible Trp rotamer; for clarity, only 30 conformations are shown.

model (FRM). In this model, a rotamer is not a single, rigid conformation of an amino acid side-chain, but rather, it is a continuous ensemble of conformations, of which the classic rigid rotamer is only the average conformation (Fig. 1). In what follows, we will refer to any one possible conformation of this ensemble as a *subrotamer*. Each flexible rotamer contains all the experimental side-chain conformations from which the rigid rotamer was derived. Moreover, with an appropriate set of rotamers for a given side-chain, the conformational space covered by the complete set of flexible rotamers will coincide with the side-chain's total conformational space.

The system of interest to side-chain modeling methods is a protein of N amino acid residues, with a rigid backbone and N flexible side-chains. Under the FRM, each side-chain has a two-level description of its conformational state: *rotamer* and *subrotamer*. The conformational state of a side-chain i can thus be represented as (r_i, s_i), where r_i denotes the rotamer and s_i the subrotamer. The total conformational state of the protein can be represented as $(\mathbf{r}, \mathbf{s})$, where $\mathbf{r} = (r_1, r_2, \ldots, r_N)$ and $\mathbf{s} = (s_1, s_2, \ldots, s_N)$.

The energy of the system in a particular conformational state $(\mathbf{r}, \mathbf{s})$ is:

$$E(\mathbf{r}, \mathbf{s}) = \sum_{i}^{N} E_i(r_i, s_i) + \tfrac{1}{2} \sum_{i}^{N} \sum_{j \neq i}^{N} E_{ij}(r_i, s_i, r_j, s_j) \qquad (1)$$

where $E_i(r_i, s_i)$ includes the intrinsic potential energy of side-chain i and its potential energy of interaction with the backbone, and $E_{ij}(r_i, s_i, r_j, s_j)$ represents the potential energy of interaction between side-chains i and j.

All thermodynamic properties of the system can be derived from its free energy:

$$A = \overline{E} - TS \qquad (2)$$

with the average energy and the entropy defined by[25]:

$$\overline{E} = \sum_{\mathbf{r}} \sum_{\mathbf{s}} p(\mathbf{r}, \mathbf{s})\, E(\mathbf{r}, \mathbf{s}) \qquad (3)$$

$$S = -R \sum_{\mathbf{r}} \sum_{\mathbf{s}} p(\mathbf{r}, \mathbf{s}) \ln p(\mathbf{r}, \mathbf{s}) \qquad (4)$$

where $p(\mathbf{r}, \mathbf{s})$ is the probability of the conformational state $(\mathbf{r}, \mathbf{s})$ and where the summations in $\mathbf{r}$ and $\mathbf{s}$ run over all possible combinations of rotamers and subrotamers of the N amino acid residues. Since the number of subrotamers belonging to a rotamer can be arbitrarily large, the summation in $\mathbf{s}$ is also arbitrarily large. In the limit of a continuous distribution, the number of subrotamers of any rotamer is infinite, as is the summation over all possible combinations of subrotamers. Therefore, in order to conserve the advantages of speed and efficiency of rotamer-based methods, it would be desirable to eliminate the explicit subrotamer level of description of the conformational state of the protein. This would correspond to

transforming the system into a thermodynamically equivalent one whose conformational states are described solely at the rotamer level. The free energy of this equivalent system is of the form:

$$A_{eq} = \overline{E}_{eq} - TS_{eq} = \sum_{\mathbf{r}} p(\mathbf{r})\, E_{eq}(\mathbf{r}) + RT \sum_{\mathbf{r}} p(\mathbf{r}) \ln p(\mathbf{r}) \quad (5)$$

and we want to find an effective energy $E_{eq}(\mathbf{r})$ such that $A_{eq} = A$.

The explicit subrotamer level of description can be eliminated from the average energy using standard concepts of probability theory, as follows:

$$\overline{E} = \sum_{\mathbf{r}} \sum_{\mathbf{s}} p(\mathbf{r}, \mathbf{s})\, E(\mathbf{r}, \mathbf{s}) = \sum_{\mathbf{r}} p(\mathbf{r}) \sum_{\mathbf{s}} p(\mathbf{s}|\mathbf{r})\, E(\mathbf{r}, \mathbf{s}) = \sum_{\mathbf{r}} p(\mathbf{r})\, \overline{E}(\mathbf{r}) \quad (6)$$

where

$$\overline{E}(\mathbf{r}) = \sum_{\mathbf{s}} p(\mathbf{s}|\mathbf{r})\, E(\mathbf{r}, \mathbf{s}) \quad (7)$$

is the average energy of the system when fixed in the rotamer configuration $\mathbf{r}$, $p(\mathbf{r})$ is the probability of this configuration and $p(\mathbf{s}|\mathbf{r})$ is the conditional probability of its sub-state $\mathbf{s}$. Similarly, for the entropy of the system:

$$\begin{aligned} S &= -R \sum_{\mathbf{r}} \sum_{\mathbf{s}} p(\mathbf{r}, \mathbf{s}) \ln p(\mathbf{r}, \mathbf{s}) \\ &= -R \sum_{\mathbf{r}} p(\mathbf{r}) \sum_{\mathbf{s}} p(\mathbf{s}|\mathbf{r}) [\ln p(\mathbf{r}) + \ln p(\mathbf{s}|\mathbf{r})] \\ &= -R \sum_{\mathbf{r}} p(\mathbf{r}) \ln p(\mathbf{r}) + \sum_{\mathbf{r}} p(\mathbf{r}) S(\mathbf{r}) \end{aligned} \quad (8)$$

where

$$S(\mathbf{r}) = -R \sum_{\mathbf{s}} p(\mathbf{s}|\mathbf{r}) \ln p(\mathbf{s}|\mathbf{r}) \quad (9)$$

is the entropy of the system when fixed in the rotamer configuration $\mathbf{r}$, due only to its underlying subrotamer disorder.

The free energy of the system can now be rewritten by inserting (6) and (8) into (2):

$$A = \sum_{\mathbf{r}} p(\mathbf{r})\, A(\mathbf{r}) + RT \sum_{\mathbf{r}} p(\mathbf{r}) \ln p(\mathbf{r}) \quad (10)$$

where

$$A(\mathbf{r}) = \overline{E}(\mathbf{r}) - TS(\mathbf{r}) \quad (11)$$

is the conditional free energy of the system when fixed in the rotamer configuration $\mathbf{r}$. Equation (10) is formally equivalent to the free energy of a system whose conformational states are described solely at the rotamer level [Eq.(5)]. Therefore, it represents the free energy of the equivalent system that we seek. Evidently, the effective energy of this system is $A(\mathbf{r})$.

Although Eq. (10) is in the required form, the calculation of $A(\mathbf{r})$ from Eqs. (7), (9), and (11) is not tractable due to the enormous number of substates $\mathbf{s}$, typical of combinatorial problems, that would have to be sampled. A way to overcome this combinatorial problem is to find a pairwise decomposition of $A(\mathbf{r})$. For the $\overline{E}(\mathbf{r})$ component of $A(\mathbf{r})$, this is always possible and straightforward. If we substitute Eq. (1) into Eq. (3) we obtain by comparison with Eq. (6):

$$\overline{E}(\mathbf{r}) = \sum_{i}^{N} \overline{E}_i(r_i) + \tfrac{1}{2} \sum_{i}^{N} \sum_{j \neq i}^{N} \overline{E}_{ij}(r_i, r_j) \quad (12)$$

where we have defined:

$$\begin{aligned} \overline{E}_i(r_i) &= \sum_{s_i} p_i(s_i|r_i)\, E_i(r_i, s_i) \\ \overline{E}_{ij}(r_i, r_j) &= \sum_{s_i} \sum_{s_j} p_{ij}(s_i, s_j|r_i, r_j)\, E_{ij}(r_i, s_i, r_j, s_j). \end{aligned} \quad (13)$$

The $S(\mathbf{r})$ component of $A(\mathbf{r})$, on the contrary, depends on the logarithm of the total conditional probability function of the system which cannot, in general, be decomposed into pairwise terms. A possible approach to obtain such a pairwise decomposition is to assume that the total probability function of the system, $p(\mathbf{r},\mathbf{s})$, and the total probability function of the rotamer configuration $\mathbf{r}$, $p(\mathbf{r})$, can be described by a generalization of Kirkwood's superposition approximation[25]:

$$\begin{aligned} p(\mathbf{r}, \mathbf{s}) &= \frac{\left[\prod_{i}^{N} \prod_{j \neq i}^{N} p_{ij}(r_i, s_i, r_j, s_j)\right]^{1/2}}{\left[\prod_{i}^{N} p_i(r_i, s_i)\right]^{N-2}} \\ p(\mathbf{r}) &= \frac{\left[\prod_{i}^{N} \prod_{j \neq i}^{N} p_{ij}(r_i, r_j)\right]^{1/2}}{\left[\prod_{i}^{N} p_i(r_i)\right]^{N-2}}. \end{aligned} \quad (14)$$

If we insert the expression for $p(\mathbf{r}, \mathbf{s})$ into Eq. (4), and take into account the expression for $p(\mathbf{r})$, we obtain by comparison with Eq. (8):

$$S(\mathbf{r}) = -(N-2) \sum_{i}^{N} S_i(r_i) + \tfrac{1}{2} \sum_{i}^{N} \sum_{j \neq i}^{N} S_{ij}(r_i, r_j) \quad (15)$$

where we have defined:

$$S_i(r_i) = -R \sum_{s_i} p_i(s_i|r_i) \ln p_i(s_i|r_i)$$

$$S_{ij}(r_i, r_j) = -R \sum_{s_i} \sum_{s_j} p_{ij}(s_i, s_j|r_i, r_j) \ln p_{ij}(s_i, s_j|r_i, r_j). \quad (16)$$

If we insert Eqs. (12) and (15) into Eq.(11) we finally obtain after some rearrangement of terms:

$$A(\mathbf{r}) = \sum_{i}^{N} A_i(r_i) + \tfrac{1}{2} \sum_{i}^{N} \sum_{j \neq i}^{N} [A_{ij}(r_i, r_j) - A_i(r_i) - A_j(r_j)] \quad (17)$$

where we have defined:

$$A_i(r_i) = \overline{E}_i(r_i) - TS_i(r_i)$$
$$A_{ij}(r_i, r_j) = [\overline{E}_i(r_i) + \overline{E}_{ij}(r_i, r_j) + \overline{E}_j(r_j)] - TS_{ij}(r_i, r_j). \quad (18)$$

$A(\mathbf{r})$ is now in the required pairwise decomposed form. $A_i(r_i)$ represents the sum of the effective intrinsic rotamer energy and the effective energy of interaction of the rotamer with the backbone whereas $\{A_{ij}(r_i, r_j) - A_i(r_i - A_j(r_j)\}$ represents the effective energy of interaction between the rotamer pair. They are analogous to the respective potential energy terms into which the total potential energy is commonly decomposed in rotamer-based methods using the rigid rotamer model (RRM) and, therefore, may be used in substitution of these.

The calculation of $A_i(r_i)$ and $A_{ij}(r_i, r_j)$ is not straightforward. Even though these quantities are of the form "energy $- T \times$ entropy," they are not true (conditional) free energies since the energetic terms $E_i(r_i, s_i)$ and $E_{ij}(r_i, s_i, r_j, s_j)$ do not completely determine the respective probabilities of interest $p_i(s_i|r_i)$ and $p_{ij}(s_i, s_j|r_i, r_j)$. On the contrary, these probabilities depend upon the complete system and are thus a function of the total energy. Therefore, their exact calculation would, once again, require the solution of the full combinatorial problem. However, if all subrotamers of a particular rotamer are sufficiently similar to one another, their conditional free energies of interaction with the rest of the system will be similar and may be considered as approximately constant. In this case, it is easy to derive the following approximate expressions for $p_i(s_i|r_i)$ and $p_{ij}(s_i, s_j|r_i, r_j)$ that depend solely on the energetic terms $E_i(r_i, s_i)$ and $E_{ij}(r_i, s_i, r_j, s_j)$:

$$p_i(s_i|r_i) \approx \exp\left[-\frac{E_i(r_i, s_i)}{RT}\right] \Big/ \sum_{s_i'} \exp\left[-\frac{E_i(r_i, s_i')}{RT}\right]$$

$$p_{ij}(s_i, s_j|r_i, r_j) \approx \exp\left[-\frac{E_i(r_i, s_i) + E_{ij}(r_i, s_i, r_j, s_j) + E_j(r_j, s_j)}{RT}\right] \Big/ \sum_{s_i'} \sum_{s_j'} \exp\left[\frac{E_i(r_i, s_i') + E_{ij}(r_i, s_i', r_j, s_j') + E_j(r_j, s_j')}{RT}\right]. \quad (19)$$

Substitution of these expressions into Eq. (13) and Eq. (16) yields the following approximate expressions for $A_i(r_i)$ and $A_{ij}(r_i, r_j)$:

$$A_i(r_i) \approx -RT \ln \sum_{s_i} \exp\left[-\frac{E_i(r_i, s_i)}{RT}\right]$$

$$A_{ij}(r_i, r_j) \approx -RT \ln \sum_{s_i} \sum_{s_j} \exp\left[-\frac{E_i(r_i, s_i) + E_{ij}(r_i, s_i, r_j, s_j) + E_j(r_j, s_j)}{RT}\right]. \quad (20)$$

The estimation of the right-hand sides of Eqs. (20) is now tractable and can be carried out by using a direct sum over states or some other method.

Technical Details

The rotamer library used in all calculations for non-Pro residue types was identical to the extended amino214 rotamer library previously described,[23] based on the amino214 rotamer library of Tufféry and co-workers.[26] Cartesian coordinates of side-chain atoms were generated from the rotamer torsion angle values as previously described.[23] For Pro, 18 rotamers were built by minimization of the acetylprolylaminomethyl dipeptide, with χ_1 either positive or negative, and φ restrained at 20°, 30°, 40°, 50°, 60°, 70°, 80°, 90°, or 100° with a harmonic potential of force constant $100RT$ at $T = 298$ K. An additional rotamer was built, also by minimization, in which all five ring atoms were held coplanar by restaining χ_1 and χ_2 with the same harmonic potential. The GROMOS package[27] was used in all minimization's and electrostatic interactions were not considered.

The calculation of the free energy terms $A_i(r_i)$ and $A_{ij}(r_i, r_j)$ was based on the following rearrangement of Eq. (17):

$$A(\mathbf{r}) = \tfrac{1}{2} \sum_{i}^{N} \sum_{j \neq i}^{N} \left[A_{ij}(r_i, r_j) - \frac{N-2}{N-1} \{A_i(r_i) + A_j(r_j)\}\right]. \quad (21)$$

For each pair of rotamers r_i and r_j, a sample of m random pairs of subrotamers was generated. The quantities $\{A_i(r_i) + A_j(r_j)\}$ and $A_{ij}(r_i, r_j)$ were then estimated by a direct sum over states based on Eqs (20):

$$\{A_i(r_i) + A_j(r_j)\} = -RT \ln \sum_{(s_i, s_j)}^{m} \exp\left[-\frac{E_i(r_i, s_i) + E_j(r_j, s_j)}{RT}\right]$$

$$A_{ij}(r_i, r_j) = -RT \ln \sum_{(s_i, s_j)}^{m} \exp\left[-\frac{E_i(r_i, s_i) + E_{ij}(r_i, s_i, r_j, s_j) + E_j(r_j, s_j)}{RT}\right] \quad (22)$$

where (s_i, s_j) represents a subrotamer pair. Sample sizes of 10, 100, and 1,000 subrotamer pairs were tested. In the present implementation of the method, the increase in computational cost for calculating these effective rotamer energies for FRM, compared to calculating the potential energies for RRM, is proportional to the size of the sample

of subrotamer pairs. Thus, for a sample of 1,000 subrotamer pairs the energy calculations with FRM are 1,000 times slower than those with RRM. However, since the calculations with RRM are on the *second* time scale, those with FRM using a sample of 1,000 subrotamer pairs are on the *hour* time scale, which is perfectly feasible.

For all non-Pro residues, subrotamers were generated by performing simultaneous random moves in torsion angles and bond angles with a Gaussian distribution centered at the rigid rotamer conformation defined in the rotamer library. Although sampling from a uniform distribution is, in principle, more correct, in several tests we found that side-chain prediction accuracy was invariably considerably lower when a uniform distribution function was used for subrotamer generation than when a Gaussian distribution was used. In spite of the physical basis for these results being unclear, they appear to indicate that better estimates of the true free energies of Eq. (18) are obtained with the approximate expressions (20) when a Gaussian as opposed to uniform distribution is used for sampling. Standard deviations for torsion angles were those obtained for the amino214 rotamer library.[28] A uniform standard deviation of 3.1° was used for all bond angles. A move was rejected if it was larger than a multiple of the respective standard deviation, σ, to avoid intersection of the conformational spaces spanned by the set of flexible rotamers of a given side-chain. Threshold values of σ, 2σ, and 3σ were tested. Aromatic rings, the -COO group of Asp and Glu, the $-CONH_2$ group of Asn and Gln and the guanidinium group of Arg were held rigid in their equilibrium conformations.

For Pro, subrotamers were generated from the respective rigid rotamer by molecular dynamics using the GROMOS package,[27] in which χ_1 and φ were restrained with a harmonic potential of $1.0RT(10°)^{-2}$ and $10.0RT(10°)^{-2}$, respectively, at $T = 298$ K.

The potential energy terms $E_i(r_i, s_i)$ and $E_{ij}(r_i, s_i, r_j, s_j)$ of Eqs. (21), and the equivalent terms for RRM, were calculated with the standard potential energy function used in GROMOS,[27] with the exception that the dielectric constant in the electrostatic term was the distance dependent dielectric constant of Solmajer and Mehler.[29,30] All parameters were from the GROMOS force field with explicit aromatic and polar hydrogen atoms.[27,31] In the calculations with the RRM, the Lennard-Jones potential between an atom pair was truncated to a maximum value of 10 kcal mol^{-1} as described by Mendes et al.[23] In all calculations the backbone and prosthetic groups were rigid and coordinates were those from the respective PDB file; when multiple conformations of the backbone existed, we used the one with the highest occupancy.

Side-chain modeling was performed with the SCMFT method[19] incorporating the MRRI protocol.[23] An update of the previously described MRRI protocol was used in which the memory coefficient λ was chosen at random in the interval [0.1,1.0]. We found that this procedure led to faster convergence and a higher population of low-lying free energy minima. With FRM, the predicted protein model was built from the average conformation of the flexible rotamer, that is, from the standard rigid rotamer. Side-chains were modeled independently with the SCWRL method[24] using the SCWRL 2.1 package.

A test set of 20 high quality proteins was used in all simulations. The PDB codes for these proteins are 2ERL, 1CBN, 5RXN, 1BPI, 1IGD, 1ISU, 1PTX, 1CTJ, 1PLC, 9RNT, 1AAC, 256B, 1WHI, 2IHL, 2END, 2HBG, 1AMM, 1XNB, 1CEX, and 1ARB. The criteria for selecting these proteins are described Mendes et al.,[23] where the same set of test proteins was used.

The predicted protein models were further refined by restrained energy minimization in vacuum using the GROMOS package.[27,31] Backbone and prosthetic group atom positions were restrained with a harmonic potential of 10^3 kJ mol^{-1} $Å^{-2}$. Side-chain torsion angles were also restrained with a harmonic potential of progressively weaker force constant the further the torsion angle is from the backbone: χ_1: $2.0RT(20°)^{-2}$, χ_2: $1.0RT(20°)^{-2}$, χ_3: $0.5RT(20°)^{-2}$, and χ_4: $0.25RT(20°)^{-2}$, at $T = 298$ K. The minimization run consisted of 500 steepest descent steps followed by 3,000 conjugate gradient steps. For models built taking electrostatics into account, an electrostatic term with a distance dependent dielectric constant[29,30] was used in the refinement; for models built neglecting electrostatics, no electrostatic term was used in the refinement.

Prediction results of all non-Pro residues were evaluated as previously described.[23] The prediction of Pro was now evaluated in the same way as other residue types except that a more restrictive cutoff of 20° was used for torsion angle correctness. ALL and CORE were as previously defined.[23]

Figures of protein structure models were generated using XtalView 3.2[32] and Raster3D 2.3.[33]

RESULTS AND DISCUSSION

Assessment of the Accuracy of Side-Chain Prediction Obtained With FRM

Comparison with RRM

To assess the effect of rotamer flexibility on the accuracy of side-chain prediction, we compared the predictions obtained from effective energies calculated with FRM with those obtained from potential energies calculated with the classic RRM, using the SCMFT method for side-chain modeling[19] incorporating the MRRI protocol[23]; all remaining conditions (for example, potential energy function, rotamer library, set of test proteins) were identical. We found that flexibility resulted in a large increase in average prediction accuracy: 7.3% in average χ_1 prediction in ALL and 6.9% in CORE, 8.1% in average χ_{1+2} prediction in ALL and 10.5% in CORE, 0.23 Å in average global RMSD in ALL and 0.43 Å in CORE (Fig. 2). Since χ_{1+2} prediction is a stricter criterion of correctness than χ_1 prediction, the large increases in average χ_{1+2} prediction and decreases in average global RMSD indicate that the quality of the protein structure models obtained with FRM is considerably higher than that of the models obtained with RRM.

The prediction of all residue types increased with flexibility and the increase for most residue types was large,

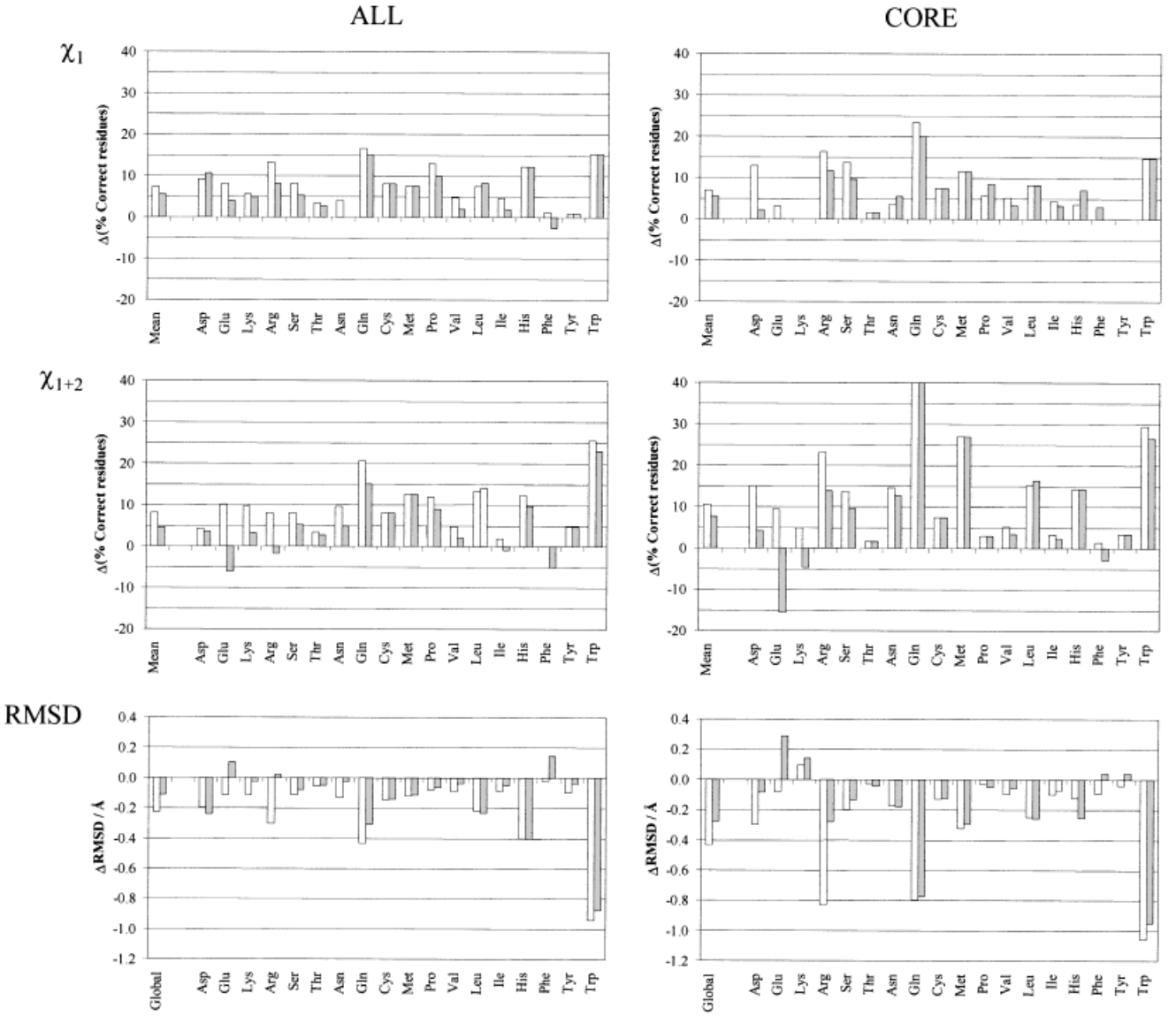

Fig. 2. Effect of rotamer flexibility on side-chain prediction accuracy. White bars: differences between the predictions obtained with FRM and those obtained with RRM. Gray bars: differences between the predictions obtained with FRM after removing the entropy terms from the effective energies [Eqs. (18)] and those obtained with RRM. Common simulation conditions for all three rotamer models: side-chain modeling method — SCMFT with the MRRI protocol; potential energy function — complete GROMOS function with a distance-dependent dielectric constant. Additional common simulation conditions for both flexible rotamer models: flexibility model — torsion angle flexibility plus bond angle flexibility; sample size: — 1,000 subrotamer pairs; threshold for subrotamer rejection — 3σ.

especially in CORE (Fig. 2). We observed the largest increases in prediction in CORE for Asp, Arg, Ser, Asn, Gln, Met, Leu, His and Trp. In Figure 3, we present an example of the prediction of the aromatic residues in the protein 1XNB, that in our test set has the largest number of Trp residues. This example shows the considerable superiority of FRM over RRM for Trp prediction and also the high quality of the protein structure models obtained with FRM.

Together, these results suggest that the effective energies calculated with FRM are a considerably more accurate measure of the energetics of a side-chain in a real protein than the potential energies calculated with RRM. This improved accuracy may be due to the energetic component of the effective energies of FRM being a more realistic description of the rotamer energy than the potential energies of RRM. Alternatively, it may result from the additional entropy term of the effective energies of FRM accounting for rotamer entropy that is nonexistent in the potential energies of RRM. Thus, to assess the relative importance of the energetic and entropic components of the effective energies to the improved description of side-chain energetics, we studied the effect of removing the entropy terms from Eqs. (18). With these incomplete

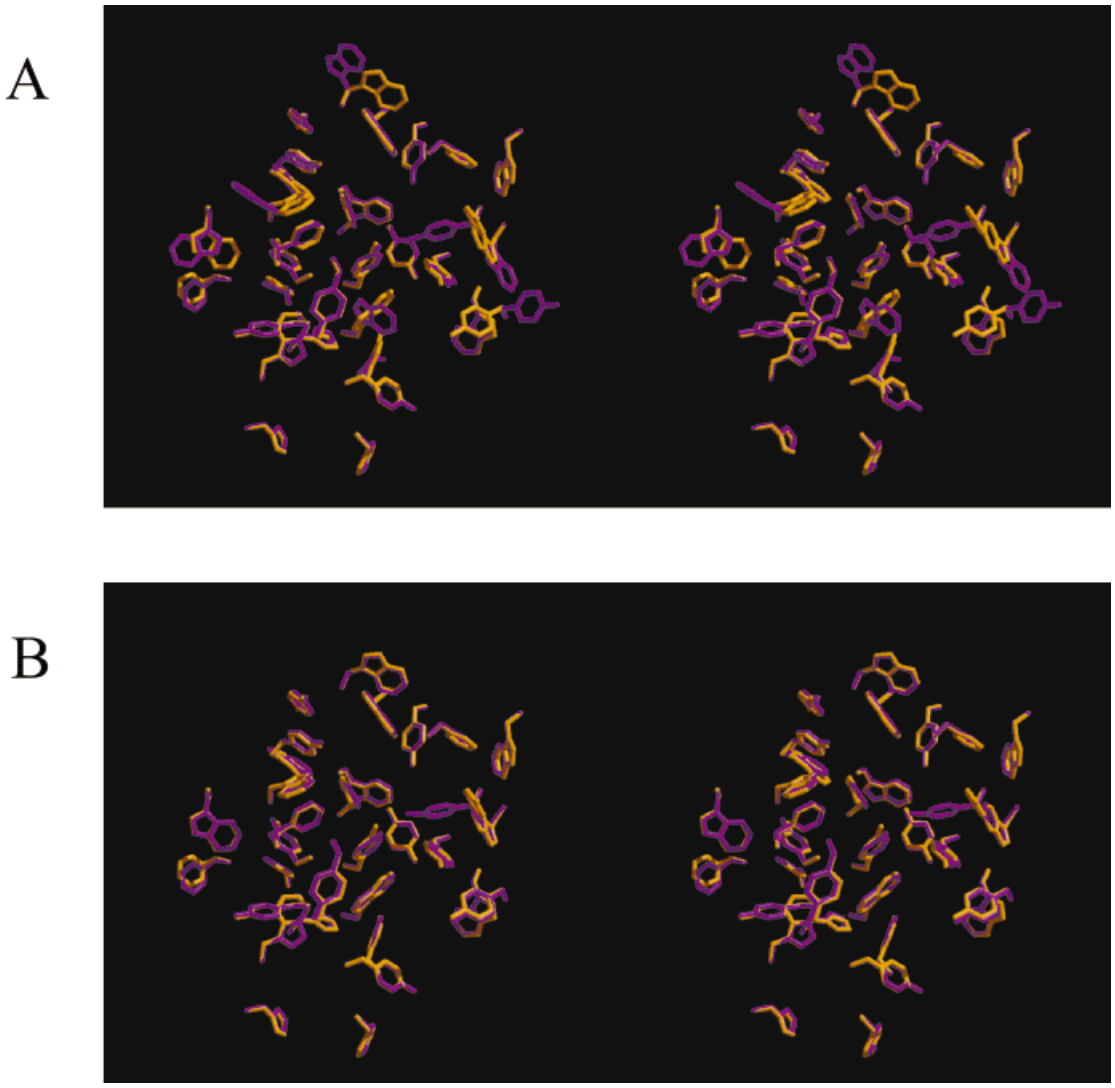

Fig. 3. Prediction of aromatic residues in the protein 1XNB: stereo view of the superposition of the protein models predicted with RRM **(A)** and with FRM **(B)** with the X-ray structure model. Yellow: X-ray structure model. Purple: predicted models. Simulation conditions for the predictions were as described in Fig. 2.

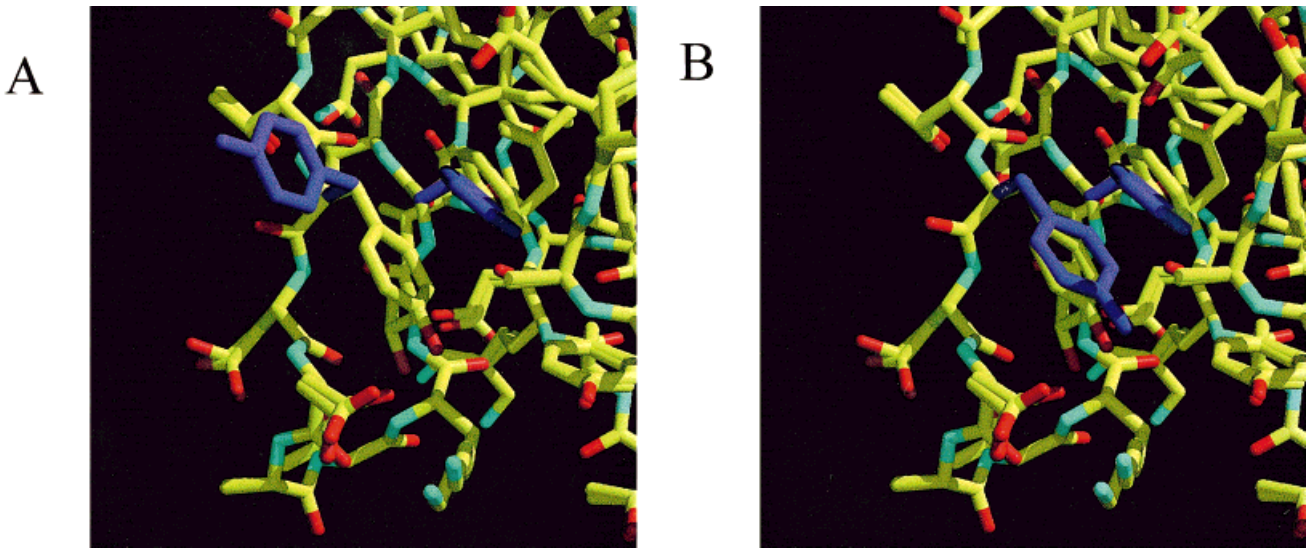

Fig. 4. Prediction of Tyr 50 and Phe 57 from the protein 1IGD: superposition of the protein models predicted with RRM **(A)** and with FRM **(B)** with the X-ray structure model. Predicted models are prior to refinement. The pair of relevant residues are colored purple for the predicted models and by atom type for the X-ray structure model. All remaining residues are not distinguished, both the predicted models and the X-ray structure model being colored by atom type. Simulation conditions for the predictions were as described in Fig. 2.

effective energies, we found that the improvement in average prediction accuracy in ALL in relation to RRM was considerably smaller than when the complete effective energies were used (Fig. 2). In particular, the increase in average χ_{1+2} prediction (4.5%) and the decrease in average global RMSD (0.11 Å) were only approximately half of that obtained with the complete energies (see above). As for residue type prediction, the improvement in prediction in ALL, especially of χ_{1+2} prediction, was smaller for almost all residue types with the incomplete effective energies, especially for residue types that are usually considerably exposed to solvent, namely, charged or polar residues.

The difference between the improvements in average prediction accuracy relative to RRM, obtained with the two effective energy functions was smaller in CORE than in ALL (Fig. 2). Moreover, the largest decreases in prediction resulting from the removal of the entropy terms were for charged residues (Asp, Glu, Lys, and Arg) that, although considered as core residues with the simple criterion of 30% or less solvent accessible surface area used here, are probably not true core residues by more detailed criteria; the improvements in prediction accuracy for all non-charged residue types was either the same or approximately the same for both energy functions. These results suggest that, for core residues, the improvement in the accuracy of side-chain energetics with the FRM effective energies is due primarily to the energetic component, the importance of the entropic component being small. In contrast, for surface residues, the energetic and the entropic components of the effective energy are of comparable importance to this improvement in accuracy. This finding is consistent with what would be expected on a physical basis: surface residues are, in general, much more flexible than core residues and, therefore, their entropy is higher and more important in determining their energetics.

If the entropic component of the effective energies of core residues in not very important, then these effective energies correspond approximately to average rotamer energies [see Eqs. (18)]. In the limiting case of severe atomic overlap, it is easy to understand how these average energies provide a more accurate measure of the rotamer's interactions than the potential energies calculated with RRM. For example, between the pair of correct rigid rotamers for Tyr 50 and Phe 57 of the core of protein 1IGD, there are ten pairs of sterically overlapping atoms (compare correct rigid rotamers of Figure 4B with corresponding experimental conformations). Using the truncated Lennard-Jones potential classically used in combination with RRM, and the truncation cutoff of 10 kcal mol^{-1} used here, the interaction between this pair of rigid rotamers will contribute on the order of 10×10 kcal mol^{-1} to the total potential energy of each rotamer. Therefore, it is obvious that a Tyr rotamer that protrudes into the solvent, and thus has no steric overlap with the correct Phe rotamer, will have a much lower potential energy of interaction with this rotamer and, therefore, will be favored in relation to the correct Tyr rotamer (Fig. 4A). In contrast, within the ensembles of subrotamers, defining the pair of correct flexible rotamers of the two residues, there are many subrotamer pairs that do not have atomic overlap, in particular, the pair of true average side-chain conformations. The average energy of interaction between the pair of correct flexible rotamers may, therefore, be favorable. Indeed, since the correct rotamer was predicted for Tyr with FRM (Fig. 4B), it is evident that this average energy of interaction was more favorable than between the correct Phe rotamer and any of the incorrect flexible Tyr rotamers.

For surface residues, such a simple interpretation as that given for core residues is not possible, since entropic effects, which are difficult to visualize, play an important role in their energetics.

Analysis of the absolute prediction accuracies obtained with FRM

The absolute prediction values obtained with the FRM are high. On average 87.7% of all core residues are predicted correctly until χ_2 and the average global RMSD for core residues is 0.70 Å (Fig. 5). The χ_{1+2} prediction of all residue types in the core, except Asp, Lys, Ser, and Pro, is above 80% and above 90% for half of these. Noteworthy is the high prediction of Trp, which in general has proven the most difficult aromatic residue to predict (see also Fig. 3).

On average, 76.5% of the residues are predicted correctly until χ_2 in ALL (Fig. 5). This lower value, and the correspondingly higher average global RMSD, compared to core residues may be due to the lower definition of surface residues in experimental protein structures. This is supported by the fact that 82.5% of the residues with an average B factor for non-hydrogen side-chain atoms below 20 Å^2, are correctly predicted to χ_2. However, the lower prediction for surface residues is also certainly due to the absence of solvation terms (hydrophobic and electrostatic) in the potential energy function we used. Such terms are much more critical in determining the energetics of surface residues than that of core residues.

Comparison with the results obtained with other methods in the literature

We compared the results obtained with FRM in combination with the MRRI SCMFT method with those obtained with the SCWRL method,[24] to our knowledge the most accurate side-chain modeling method in the literature. To allow a direct comparison of the two methods, we applied SCWRL to our set of test proteins. We found that the average predictions obtained with FRM (Fig. 5) were higher than those obtained with SCWRL. Although significant, the differences in ALL were not very large: 2.4% in average χ_1 prediction, 4.2% in average χ_{1+2} prediction and 0.14 Å in average global RMSD. In CORE, on the other hand, the differences were considerable: 5.3% in average χ_1 prediction, 8.7% in average χ_{1+2} prediction and 0.48 Å in average global RMSD. The discrepancy between ALL and CORE is due to the fact that the average prediction obtained by SCWRL for surface residues is approximately 1% better both for χ_1 and for χ_{1+2} than that obtained with FRM (results not shown), whereby the higher prediction in ALL with FRM is due only to the better prediction of core

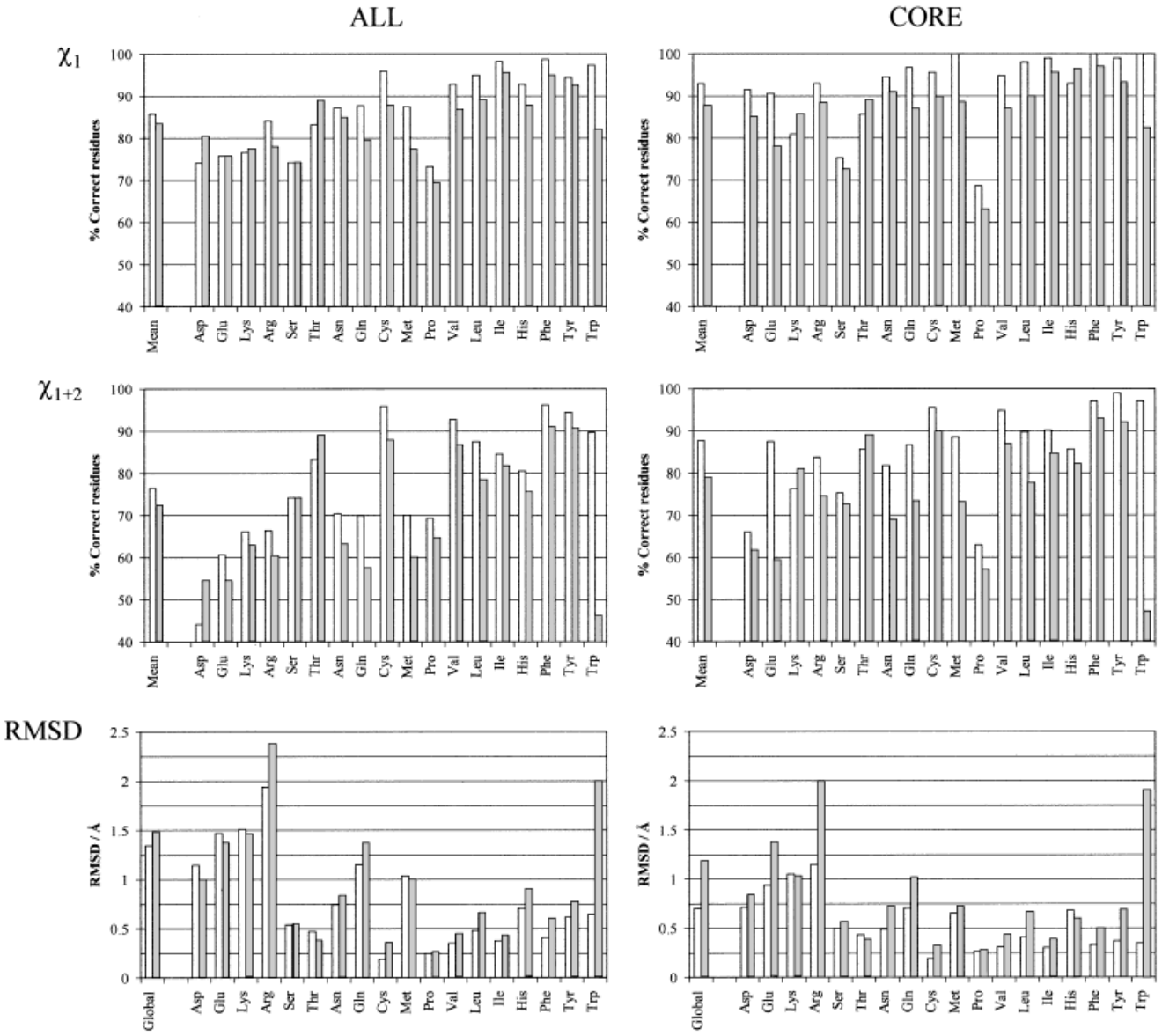

Fig. 5. Absolute side-chain predictions obtained with FRM (white bars) and with SCWRL (gray bars). Simulation conditions for the predictions with FRM were as described in Fig. 2.

residues. These findings are consistent with two facts. On the one hand, problems of atomic overlap should be more severe for core residues than for surface residues and, thus, FRM should perform better for these residues than the RRM used in SCWRL. On the other hand, the potential energy function used with FRM did not contain solvation terms that are critical in determining the energetics of surface residues and, therefore, a high prediction accuracy cannot be expected for these residues.

In relation to the prediction of individual residue types, we found that, with the exception of Asp, Ser, and Thr in ALL and Lys and Thr in CORE, the χ_{1+2} prediction of all other residue types was higher with FRM, the differences being considerable for most residue types, especially in CORE. The most noticeable example is Trp whose χ_{1+2} prediction with FRM was more than double that of SCWRL.

Analysis of the Influence on Side-Chain Prediction Accuracy of Several Factors Upon Which Calculation of Effective Energies With FRM Depends

We have analyzed the influence on side-chain prediction accuracy of several factors upon which the effective energy calculations with FRM depend. These factors included the fraction of the full conformational space of the flexible rotamer that is sampled, the size of the sample of random subrotamer pairs used to estimate the effective energies, the removal of flexibility from bond angles, and the

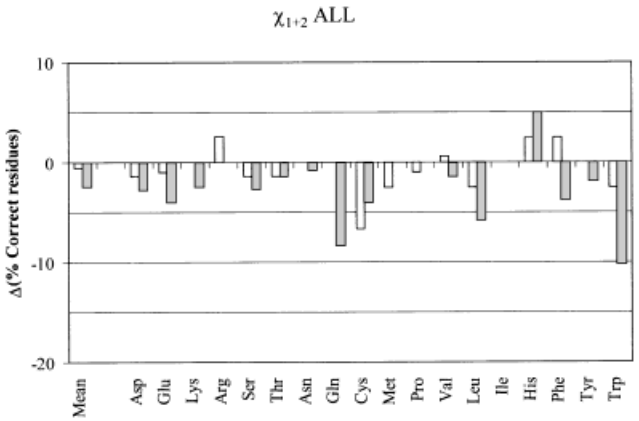

Fig. 6. Effect on side-chain prediction accuracy with FRM of the fraction of rotamer conformational space sampled in the calculation of effective energies: differences between the predictions obtained with a subrotamer rejection threshold of 2σ (white bars) or σ (gray bars) and those obtained with a threshold of 3σ. Common simulation conditions for all three threshold values: side-chain modeling method — SCMFT with the MRRI protocol; potential energy function — complete GROMOS function with a distance-dependent dielectric constant; flexibility model — torsion angle flexibility plus bond angle flexibility; sample size —1,000 subrotamer pairs.

Fig. 7. Effect on side-chain prediction accuracy with FRM of the size of the sample of subrotamer pairs used in the calculation of effective energies: differences between the predictions obtained with a sample size of 100 subrotamer pairs (white bars) or 10 subrotamer pairs (gray bars) and those obtained with a sample size of 1,000 subrotamer pairs. Common simulation conditions for all three sample sizes: side-chain modeling method — SCMFT with the MRRI protocol; potential energy function — complete GROMOS function with a distance-dependent dielectric constant; flexibility model — torsion angle flexibility plus bond angle flexibility; threshold for subrotamer rejection — 3σ.

removal of the electrostatic term from the potential energy function. In all of these studies, we will only present the results for χ_{1+2} prediction in ALL (and also for RMSD prediction in ALL in one case), since the information contained in all other results was redundant.

Fraction of conformational space of flexible rotamer sampled in effective energy calculation

The fraction of the conformational space of the flexible rotamer sampled in the calculation of effective energies was controlled by rejecting subrotamers with torsion angles or bond angles that deviated from the average conformation (rigid rotamer) by more than a multiple of the respective standard deviation, σ. We compared threshold values of 3σ, 2σ, and σ that correspond to 99.7%, 95.4% and 68.3% of the conformational space of the rotamer, respectively. From another point of view, these values represent the fraction of the original experimental side-chain conformations from which the rigid rotamer was built, that are in fact sampled in the energy calculation. We observed a small decrease in average prediction for 2σ, and a considerably larger decrease for σ, relative to 3σ (Fig. 6). The prediction of many residue types decreased for 2σ in relation to 3σ, although an increase in prediction was seen for a small number of residues. However, for σ, the prediction of almost all residue types decreased relative to 3σ and only that of His increased. Although the decrease in average prediction of either 2σ or σ relative to 3σ is not very large, the decrease in prediction for some residue types was substantial.

On a physical basis, the more completely the conformational space of a rotamer is sampled, the more accurate its effective energies should be, independent of residue type. The progressive decrease in prediction accuracy from 3σ to 2σ to σ, and the decreases in prediction observed for almost all residue types for σ relative to 3σ, are thus in accordance with what physically would be expected.

Size of sample of random subrotamer pairs used in energy calculation

We compared the side-chain predictions obtained with effective energies estimated from samples of 1,000, 100 and 10 random subrotamer pairs. We found that there was a progressive decrease in average prediction accuracy with the decrease in sample size (Fig. 7). The same applied to the decreases in the prediction of individual residue types. Although the decrease in average prediction for the sample of 100 pairs in relation to the sample of 1,000 pairs is not very large, the decrease for the sample of 10 pairs is considerable. However, even for the sample of 100 pairs, the decrease in prediction for particular residue types such as Trp is considerable.

On a physical basis, the larger the size of the sample of subrotamer pairs from which the effective energies are estimated, the more accurate these effective energies should be, independent of residue type. The progressive decrease in average prediction accuracy with the decrease in sample size, and the decreases in prediction observed for almost all residue types for the sample of 10 pairs relative to that of 1,000 pairs, are thus in accordance with what physically would be expected.

The use of a sample size larger than 1,000 pairs would probably improve prediction further. However, based on the smaller difference in prediction accuracy between the sample sizes of 1,000 pairs and 100 pairs than between the sample sizes of 100 pairs an 10 pairs, an increase in sample size to 10,000 pairs would probably only produce a small additional increase in prediction. Therefore, the

Fig. 8. Effect on side-chain prediction accuracy with FRM of removing bond angle flexibility from FRM: differences between the predictions obtained with only torsion angle flexibility and those obtained with both torsion angle and bond angle flexibility. Common simulation conditions for both flexibility models: side-chain modeling method — SCMFT with the MRRI protocol; potential energy function — complete GROMOS function with a distance-dependent dielectric constant; sample size — 1,000 subrotamer pairs; threshold for subrotamer rejection — 3σ.

sample size of 1,000 pairs appears to be a good compromise between prediction accuracy and speed of calculation.

Removal of flexibility from bond angles

We have compared the prediction accuracy obtained using a FRM with flexibility both in torsion angles and in bond angles with that obtained using a FRM with flexibility only in torsion angles. We found a significant decrease in average prediction and a decrease in the prediction of most residue types when flexibility in bond angles was removed (Fig. 8), suggesting that flexibility in bond angles is important in determining the effective energies of most residue types. The very large decrease in prediction for Trp shows that it is particularly important for this residue. This can be explained by the fact that Trp has a large number of atoms in a rigid planar group whose positions are not independent of each other and, therefore, optimal packing of this group, especially in the protein core, requires not only adjustments of the two torsion angles of this residue but also of its bond angles. This hypothesis is consistent with the fact that, for many of the Trp residues in the set of test proteins that we used, the experimental value of the $C_\alpha C_\beta C_\gamma$ angle deviated considerably from the corresponding equilibrium value of the GROMOS force field.

Removal of electrostatic term from potential energy function

We compared the side-chain prediction accuracy obtained with the complete potential energy function comprising an electrostatic term with that obtained with a potential energy function in which such a term was absent. We found a considerable decrease in average prediction when the electrostatic term was removed from the potential energy function (Fig. 9). We also found that the prediction

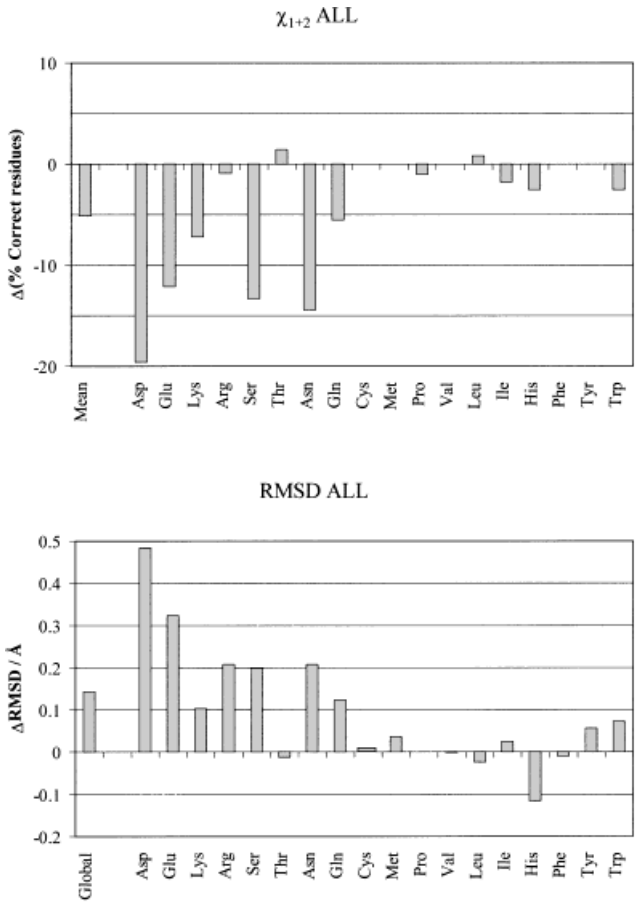

Fig. 9. Effect on side-chain prediction accuracy with FRM of removing the electrostatic term from the potential energy function: differences between the predictions obtained in the absence of the electrostatic term and those obtained in the presence of this term. Common simulation conditions for both potential energy functions: side-chain modeling method — SCMFT with the MRRI protocol; flexibility model — torsion angle flexibility plus bond angle flexibility; sample size — 1,000 subrotamer pairs; threshold for subrotamer rejection — 3σ.

of all charged or polar residue types, except Thr, decreased and that the decrease was very large for most of these residue types (the relatively large decrease in RMSD for His is probably not significant compared to the decrease in χ_{1+2} prediction, since it is due only to a small increase in the less important χ_1 prediction). The prediction of the majority of nonpolar residue types either remained unaltered or varied comparatively very little. These results are in accordance with what would be expected on a physical basis. It is also interesting to note that, although the χ_{1+2} prediction of Arg does not decrease significantly, there is a large increase in its average RMSD, suggesting that the prediction of torsion angles beyond χ_2 is considerably better with the electrostatic potential, probably due to a better positioning of the guanidinium group of the residue. A similar situation was observed for Lys in CORE (results not shown). Although the decreases in prediction for the aromatic residues His and Trp may be due to the indirect effect of altered packing of nearby charged or polar resi-

dues, as is certainly the case for the nonpolar residues Pro and Ile, the dipole and quadrupole moments on the aromatic groups of these residues may also be responsible for the better prediction in the presence of an electrostatic potential.

These results show that, contrary to what has been described in the literature for RRM (for example[23,24]), with FRM the introduction of an electrostatic term in the potential energy function does improve the prediction of practically all polar or charged residues significantly. The fact that this was previously not observed is probably derived from the classic truncation of the Lennard-Jones potential in energy calculations with RRM that may lead to an imbalance between the electrostatic and van der Waals energy terms and, thus, large errors in the total potential energy. Since the Lennard-Jones potential is not truncated in the effective energy calculations with FRM, such an imbalance does not exist. Furthermore, and perhaps more importantly, since polar and charged residues exist mainly at the protein surface, the entropic component of their effective energies is extremely important in determining their conformation, as shown above. In RRM, this component is absent and, therefore, the rotamer energies are subject to a large error even if an electrostatic term is introduced.

CONCLUSIONS

Modeling the side-chain conformations of amino acid residues in protein molecules from the residue sequence and a model of the protein backbone is a natural subproblem of many current efforts in protein tertiary structure determination and prediction. Of the existing side-chain modeling methods, rotamer-based methods are the fastest and most efficient. We have presented a flexible rotamer model (FRM) and a thermodynamically based method for calculating effective rotamer energies with this model. This method is completely general and the effective energies it provides can be used in combination with any of the rotamer-based side-chain modeling methods commonly used.

We have shown that the effective energies calculated with FRM result in a large increase in the accuracy of side-chain prediction relative to that obtained with potential energies calculated with the classic rigid rotamer model (RRM), when the same side-chain modeling method is used with both. We have also shown that the prediction accuracy obtained with FRM in combination with the MRRI SCMFT side-chain modeling method is significantly higher than that of other methods in the literature, especially for core residues.

The high side-chain prediction accuracies obtained with the FRM indicate that it can be used with considerable advantage in all areas of protein research where side-chain modeling is required. Thus, side-chain modeling methods incorporating FRM can be used in de novo protein design methods, as nonbiased methods for the initial placement of side-chains in protein structure determination by x-ray crystallography and as a tool for side-chain placement in theoretical methods for predicting protein tertiary structure, such as homology-based modeling or threading algorithms. Preliminary work on homology-based modeling has already been undertaken.[34,35]

The results obtained here also indicate that the effective energies calculated with FRM are a considerably more accurate measure of the side-chain energetics in a real protein than the potential energies calculated with the classic RRM. This suggests the use of FRM in free energy calculations, such as the free energies of mutation in mutational studies.

ACKNOWLEDGMENTS

The authors are very grateful to Dr. Pierre Tufféry for supplying the amino214 rotamer library and the respective standard deviations in torsion angles. They also thank Dr. Roland Dunbrack and Dr. Michael Bower for supplying the SCWRL programs and for invaluable help with running them and Dr. Luis Serrano for helpful discussions.

REFERENCES

1. Martin ACR, MacArthur MW, Thornton JM. Assessment of comparative modeling in CASP2. Proteins 1997;Suppl. 1:14–28.
2. Marchler-Bauer A, Bryant SH. Measures of threading specificity and accuracy. Proteins 1997;Suppl. 1:74–82.
3. Zemla A, Venclovas C, Reinhardt A, Fidelis K, Hubbard T. Numerical criteria for the evaluation of ab initio predictions of protein structure. Proteins 1997;Suppl. 1:140–150.
4. Kono H, Doi J. Energy minimization method using automata network for sequence and side-chain conformation prediction from given backbone geometry. Proteins 1994;19:244–255.
5. Dahiyat BI, Mayo SL. De novo protein design: fully automated sequence selection. Science 1997;278:82–87.
6. Kortemme T, Ramirez-Alvarado M, Serrano L. Design of a 20-amino acid, three-stranded beta-sheet protein. Science 1998;281: 253–256.
7. Kolinski A, Skolnick J. Monte Carlo simulations of protein folding. I. Lattice model and interaction scheme. Proteins 1994;18;338–352.
8. Vásquez M. Modeling side-chain conformation. Curr Opin Struct Biol 1996;6:217–221.
9. Janin J, Wodak S, Levitt M, Maigret B. Conformation of amino acid side-chains in proteins. J Mol Biol 1978:125:357–386.
10. Gelin BR, Karplus M. Side-chain torsional potentials: effect of dipeptide, protein and solvent environment. Biochemistry 1979;18: 1256–1268.
11. Schrauber H, Eisenhaber F, Argos P. Rotamers: to be or not to be? An analysis of amino acid side-chain conformations in globular proteins. J Mol Biol 1993;230:592–612.
12. Roitberg A, Elber R. Modeling side-chains in peptides and proteins: application of the locally enhanced sampling and simulated annealing methods to find minimum energy conformations. J Chem Phys 1991;95:9277–9287.
13. Lee C, Subbiah S. Prediction of protein side-chain conformation by packing optimization. J Mol Biol 1991;217:373–388.
14. Lee C. Predicting protein mutant energetics by self-consistent ensemble optimization. J Mol Biol 1994;236:918–939.
15. McCammon JA, Harvey SC. Dynamics of proteins and nucleic acids. Cambridge: Cambridge University Press; 1987.
16. Brook CLI, Karplus M, Pettit BM. Proteins: a theoretical perspective of dynamics, structure and thermodynamics. New York: John Wiley & Sons; 1987.
17. Levitt M. Accurate modelling of protein conformation by automatic segment matching. J Mol Biol 1992;226:507–533.
18. Holm L, Sander C. Fast and simple Monte Carlo algorithm for side-chain optimization in proteins: application to model building by homology. Proteins 1992;14:213–223.
19. Koehl P, Delarue M. Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. J Mol Biol 1994;239:249–275.

20. Vásquez M. An evaluation of discrete and continuum search techniques for conformational analysis of side-chains in proteins. Biopolymers 1995;36:53–70.
21. Lazar GA, Desjarlais JR, Handel TM. De novo design of the hydrophobic core of ubiquitin. Prot Sci 1997;6:1167–1178.
22. Desmet J, de Maeyer M, Hazes B, Lasters I. The dead-end elimination theorem and its use in protein side-chain positioning. Nature 1992;356:539–542.
23. Mendes J, Soares CM, Carrondo MA. Improvement of side-chain modelling in proteins with the self-consistent mean field theory method based on an analysis of the factors influencing prediction. Biopolymers 1999;50:111–131.
24. Bower MJ, Cohen FE, Dunbrack RLJ. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. J Mol Biol 1997;267:1268–1282.
25. Hill TL. Statistical mechanics. New York: McGraw-Hill; 1956.
26. Tufféry P, Etchebest C, Hazout S. Prediction of protein side-chain conformations: a study on the influence of backbone accuracy on conformation stability in the rotamer space. Protein Eng 1997;10: 361–372.
27. van Gunsteren WF, Berendsen HJC. Groningen molecular simulation (GROMOS) library manual. Groningen, The Netherlands: Biomos B. V., Biomolecular Software; 1987.
28. Tufféry P, Etchebest C, Hazout S. http://condor.urbb.jussieu.fr/rotamer-stdv.html; 1997.
29. Solmajer T, Mehler EL. Electrostatic screening in molecular dynamics simulations. Protein Eng 1991;4:911–917.
30. Solmajer T, Mehler EL. Modeling solvent effects on molecular dynamics simulations of proteins. Int J Quant Chem 1992;44:291–299.
31. Smith LJ, Mark AE, Dobson CM, van Gunsteren WF. Comparison of MD simulations and NMR experiments for hen lysozyme: analysis of local fluctuations, cooperative motions and global changes. Biochemistry 1995;34:10918–10931.
32. McRee DE. A visual protein crystallographic software system for X11/XView. J Mol Graph 1992;10:44–46.
33. Merritt EA, Bacon DJ. Raster3D: photorealistic molecular graphics. Meth Enzymol. 1997;277:505–524.
34. Burke DF, Deane CM, Nagarajaram HA, et al. An iterative structure-assisted approach to sequence alignment and comparative modelling. Proteins 1999; Suppl 3:55–60.
35. Pereira MM, Santana M, Soares CM, et al. The caa_3 terminal oxidase of the thermohalophilic bacterium *Rhodothermus marinus:* characterisation of a HiPIP:oxygen oxidoreductase lacking the key glutamate of the D-channel. BBA 1999;1413:1–3.