CODEN IEJMAT *Internet Electronic Journal of Molecular Design* **2005**, *4*, 659–670 ISSN 1538–6414
BioChem Press http://www.biochempress.com

Inter*net* Electronic Journal of Molecular Design

September 2005, Volume 4, Number 9, Pages 659–670

Editor: Ovidiu Ivanciuc

Proceedings of the Internet Electronic Conference of Molecular Design 2004
IECMD 2004, November 29 – December 12, 2004

Structure Prediction of Segments with Low Target–Template Similarity in Comparative Protein Modeling Using a Reduced Protein Model

Andrzej Szymoszek [1] and Martin Zacharias [2]

[1] Institute of Molecular Biotechnology, Beutenbergstr. 11, 07745 Jena, Germany
[2] School of Engineering and Science, International University Bremen, Campus Ring 1, 28759 Bremen, Germany

Received: November 24, 2004; Revised: May 30, 2005; Accepted: June 15, 2005; Published: September 30, 2005

Citation of the article:
A. Szymoszek and M. Zacharias, Structure Prediction of Segments with Low Target–Template Similarity in Comparative Protein Modeling Using a Reduced Protein Model, *Internet Electron. J. Mol. Des.* **2005**, *4*, 659–670, http://www.biochempress.com.

Inter*net* Electronic Journal of Molecular Design

BioChem Press
http://www.biochempress.com

Structure Prediction of Segments with Low Target–Template Similarity in Comparative Protein Modeling Using a Reduced Protein Model

Andrzej Szymoszek [1,*] and Martin Zacharias [2]

[1] Institute of Molecular Biotechnology, Beutenbergstr. 11, 07745 Jena, Germany
[2] School of Engineering and Science, International University Bremen, Campus Ring 1, 28759 Bremen, Germany

Received: November 24, 2004; Revised: May 30, 2005; Accepted: June 15, 2005; Published: September 30, 2005

Internet Electron. J. Mol. Des. **2005**, *4* (9), 659–670

Abstract

Motivation. Theoretical prediction of protein structures is important because the number of sequenced proteins grows much faster than the number of experimentally determined 3D structures. Among theoretical methods, homology or comparative modeling of unknown 3D protein structures (targets) has been established. It is based on experimental structures of proteins (templates) with sequence similarity to the target. The method is, however, limited by the degree of sequence identity. Frequently, the target– template sequence alignment is non–uniform along the sequence. In the present study the possibility to model segments of low target–template similarity by a systematic conformational search based on a reduced protein model has been explored. The force field is based on the concept of residue– residue contact energies and allows to generate a large number of putative conformations by energy minimization and selection of favorable conformations. The approach was tested on a protein of known structure by splitting the protein into mobile and conformationally restrained regions. The mobile regions represented putative regions with no structural information from a template (the conformationally restrained regions represented segments that can be modeled accurately based on a template). The residue–based reduced protein model does not allow accurate structure prediction of a complete protein. However, our results demonstrate that with the test system and the present method it is possible to successfully pick out segment topologies close to experiment among a variety of possible structures, if the rest of the protein structure is accurately defined. The approach could be useful in comparative modeling in cases where most of the target protein can be modeled accurately except for segments (beyond the length of a loop) for which no template structure is available.

Method. Randomly generated protein segment structures are subjected to energy minimization employing a reduced protein model and using positional restraints for conserved parts of the protein structure as well as distance constraints to enforce a preset secondary structure. The alpha–helical test protein results are compared to the experimental protein structure.

Results. There is a correlation between energy of a reduced protein structure, and its similarity to the experimentally known structure, evaluated by the root mean square deviation (rmsd) of corresponding atoms. Low energy structures can be pre–selected for further refinement.

Presented in part at the Internet Electronic Conference of Molecular Design 2004, IECMD 2004.

* Correspondence author; present address: UFZ–Center for Environmental Research, Leipzig, Germany; phone: +49–341–325–2031; fax: +49–341–325–2401; E-mail: andrzej.szymoszek@ufz.de.

Conclusions. Our reduced protein modeling approach has been developed as a possible tool to improve homology modeling in regions of low target– template sequence similarity. Although the initial tests of the model on a mainly alpha–helical structure showed quite reasonable performance, further testing of the model is required to make this approach generally applicable.

Keywords. Homology modeling; reduced protein model; energy minimization; residue–residue contact energies.

1 INTRODUCTION

The formation of the three–dimensional structure of a protein from a given sequence of amino acids is one of the most important problems in molecular biology, in particular in molecular modeling. There are two basic theoretical approaches: *ab initio* methods have been developed in order to predict the 3D protein structure from scratch, whereas comparative modeling of protein structures is based on recognition of homology (in most cases sequence similarity) to a template of an experimentally known structure. The latter method is limited by the degree of the target–template sequence identity. Frequently, the quality of the target– template sequence alignment is non–uniform along the sequence: parts can be modeled with a high confidence, whereas other parts differ strongly from the template. Segments of the target sequence that have no equivalent regions in the template structure (insertions or loops) are the most difficult regions to model [1]. They are often larger than small loop segments. Since at atomic resolution the accurate loop prediction is limited to short loops of up to 9 residues [1,2], one needs to evaluate a large number of possible conformations.

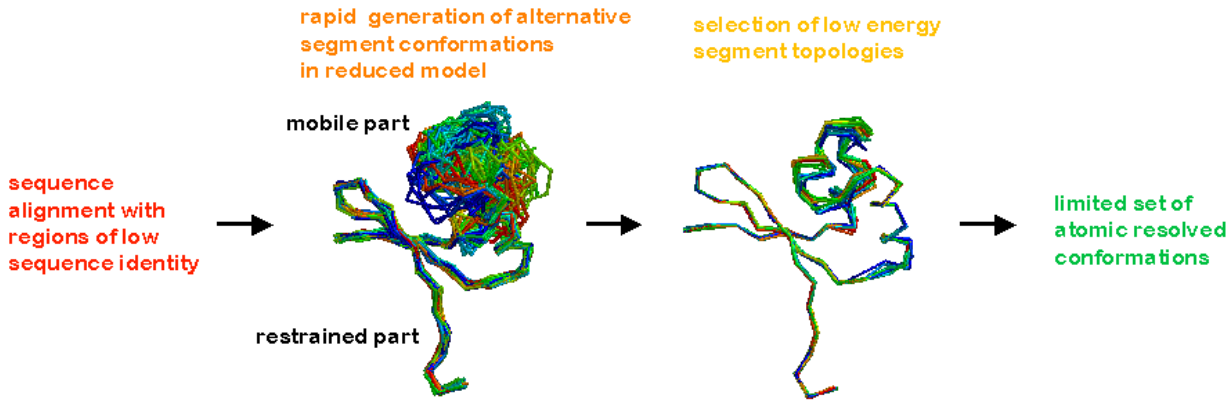


Figure 1. Schematic presentation of the reduced model approach to homology modeling.

For pre–selection of possible protein segment 3D topologies, we propose an application of a reduced protein model. It allows a very rapid generation of protein segment conformations, compatible with the boundaries imposed by those parts of the protein chain, that can be accurately modeled based on the template structure. In contrast to threading based fold recognition approaches, the present method allows in principle the generation of partially new topologies that are derivatives

of existing protein fold topologies. The idea is schematically presented in Figure 1, and outlined in the following.

A number of reduced protein models have been developed over the years, in most cases meaningful only for particular systems, or particular properties. Levitt and Warshell in middle 70ties simplified conformational energy calculations with the use of the concept of time– averaged forces [3]. The peptide group, forming the protein backbone repeating unit, was simplified by combining atoms into two effective nitrogen and oxygen pseudo atoms. Side chains were treated as single pseudo atoms located at the centroid of the side chain atoms (including backbone C–α's). The model worked well in representing a stable near native conformation of pancreatic trypsin inhibitor, and in a limited number of other cases. Honeycutt and Thirumalai developed a model to illustrate the metastability hypothesis [4]. Amino acid residues were divided into three groups: hydrophobic, hydrophilic and neutral. Long– range potentials (with an attractive term only for interactions between hydrophobic residues), bond–angle potentials and torsional potentials were included in the simplified force field, and the whole approach was restricted to beta–barrel conformations only. Wallqvist and Ullner also used a hydrophobicity criterion to parameterize interactions in their model, containing per amino acid one backbone interaction site, and depending on the size and complexity, one or two side chain sites [5]. Liwo *et al*. described united–residue potentials for off–lattice simulations, where the chain was represented by a sequence of C–α's, linked by virtual bonds with attached united side chains and united peptide groups [6]. A distance geometry approach, using a small number of distance constraints, was applied to a global fold determination by Aszodi et al. in their program DRAGON–2 [7]. Only one site per amino acid was sufficient for description of interactions in the model by Ulrich *et al*. [8]; sites interacted in one of four ways depending on their topological proximity. Another more sophisticated approach was used by Smith and Hall [9] that included four beads, three for a backbone and one for a side chain. The results of reduced protein model simulation studies have helped to better understand the principles of protein folding. However, even in favorable cases predicted structures of complete proteins based on a reduced model often differ from experiment considerably (several angstroms).

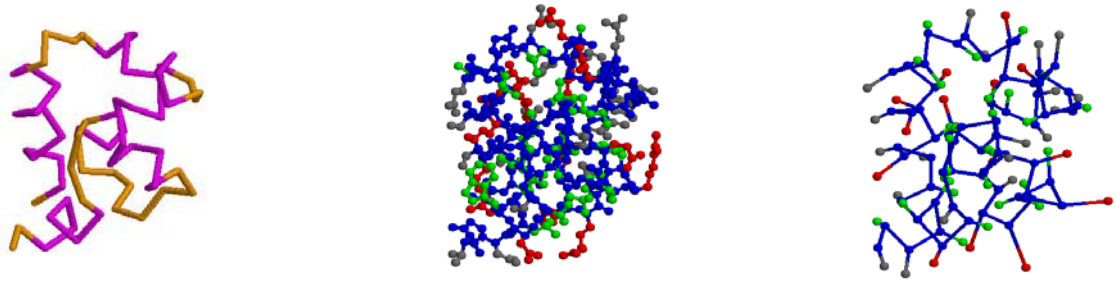
It is important to emphasize that in the present study we do not intend to perform structure prediction of complete proteins using a reduced protein model. Instead, aim of the study is to test the possibility to use a reduced protein model to predict the structure of segments of proteins assuming that the rest of the protein has been modeled accurately (based for example on high sequence similarity to a template structure). This problem is less difficult than predicting the complete structure of a protein based on a reduced model since the accurately modeled part of the protein acts as conformational constraint for the mobile segment. Since some segments are, indeed, predicted from scratch, the present method can be thought of as a kind of bridge, joining the two basic approaches of theoretical protein structure determination (homology modeling and *ab initio* protein modeling). To our knowledge, this approach is unique and therefore results cannot be

directly compared to those obtained with the use of other reduced protein models. A specific new feature of the reside–residue interaction potential of the present reduced model is that it employs a Lennard–Jones type function with an intermediate flat regime of constant residue–residue interaction (see Methods). This feature makes the calculated effective interaction energy near the optimum residue–residue distance robust with respect to small deviations from the optimal distance compared to a continuously changing potential.

As a modeling test system the amino terminal domain of phage 434 repressor, PDB– entry 1R69 [10] was chosen. It contains 63 amino acid residues with 5 alpha helices joined by loops (Figure 2a). Pairs of consecutive helices (1–2, 2–3, 3–4, 4–5) including loops between them are our mobile segments for test purposes, which leads to 4 tests in total. In each case, the rest of the protein was restrained to experimental positions, so the protein mobility pattern can be described as R(estrained)–M(obile)–R(estrained) in each of the four cases. Several hundreds of energy–minimized conformations were generated for each case. On the basis of the reduced model energy function the favorable conformations were generally relatively close in rmsd to the experimental structure.

2 MATERIALS AND METHODS

The phage 434 repressor (further referred to as 1R69) was chosen as a test example to evaluate the reduced model performance in segments containing alpha helices and one loop. Although the chain is relatively short (63 residues), it contains 16 of 20 amino acids types (exceptions are CYS, HIS, MET and TYR; parameters for those residues are given in the force field description as well). Our test protein is presented in Figure 2a–c. The alpha– helical type of 1R69 can be recognized from Figure 2a. In Figure 2b–c the difference between atomic resolution (2b) and reduced representation (2c) is demonstrated.



(*a*) (*b*) (*c*)

Figure 2. Crystal structure of 1R69: (*a*) backbone with helical segments indicated magenta, (*b*) atomic resolution, (*c*) reduced representation. In (*b*) and (*c*) protein backbone is in blue, hydrophobic side–chains in green, charged side chains in red, others in gray.

The protocol for reduced structure generation and evaluation is as follows:

1. For a given protein sequence the topology file is generated. Parameters for particular residues depend on the residue type and the known (as in our test case) or predicted secondary structure of the protein.

2. An initial structure, based on an atomic resolution PDB (protein data bank) file is prepared. This state can be either a known experimental structure (for text cases as in the present study) or a preliminary homology model (obtained from homology modeling program such as MODELLER [11]) or a structure with no coordinates for a given segment. For each residue, positions of two pseudo atoms, CA and CB, are generated. CA positions are simply original C–α atom coordinates. The equilibrium CB atom positions for each residue are given by the average distance of the center of geometry of each side chain with respect to the CA position of the residue (r_B stored in the topology file and given in Table 1).

Table 1. Residue type–dependent force field parameters

residue type	r_B[nm]	$r^e_{1/2}$[nm]
ALA	0.1621	0.1964
ARG	0.4824	0.3535
ASN	0.2616	0.2719
ASP	0.2579	0.2732
CYS	0.2328	0.2438
GLN	0.3528	0.3024
GLU	0.3535	0.3015
GLY	0.1000	0.1710
HIS	0.3151	0.3011
ILE	0.2558	0.3023
LEU	0.2715	0.2906
LYS	0.3945	0.3272
MET	0.3589	0.2940
PHE	0.3468	0.3212
PRO	0.1929	0.2746
SER	0.1998	0.2413
THR	0.2104	0.2723
TRP	0.3804	0.3481
TYR	0.3936	0.3389
VAL	0.2125	0.2788

3. The next step is the division of the protein into segments that will be treated as restrained (R) or mobile (M). As mentioned above, experiences so far are based on R–M–R structure scheme, but in principle the number of M's separated by R's can be greater than 1. Such cases will be investigated in the future. According to the R–M–R division, restraint data file for R segments is prepared. If the M segment contains regular secondary structures (alpha helices or beta strands), distance constraints are additionally prepared, so that this conformation could be retained during energy minimization.

4. The pseudo atoms of the M segment are randomly placed initially and the energy minimization of the whole structure follows. Finally obtained energy minima are subjected to evaluation, according to the calculated total energy.

The energy minimization is performed using the conjugate gradient algorithm. The interactions are defined for each of the terms in the following expression for the total energy:

$$E_{tot} = E_{bonds} + E_{bondangles} + E_{torsions} + E_{impropers} + E_{nonbonded} + E_{restr} + E_{constr} \qquad (1)$$

The first three terms have the standard form of molecular mechanics force fields with quadratic bond length and bond angle terms between consecutive pseudo atoms of the chain and cosine terms to describe the dihedral angle energy for the reduced model chain. The parameters are based on the statistical evaluation of experimental protein structures. In addition an improper dihedral between three consecutive CA atoms and a CB pseudo side chain atom was used to control the chirality of the side chain placement. The bonded interactions provide the integrity of a reduced chain representation. In folded structures contacts between residues close to each other in space are of special importance. They are described by pairwise non–bonded interactions:

$$E_{nonbonded} = \sum_{i<j} E_{CAiCAj} + E_{CAiCBj} + E_{CBiCAj} + E_{CBiCBj} \qquad (2)$$

"CA" terms in (2) are residue type independent and are given by a soft van der Waals type 6–8 expression:

$$E_{ij} = \begin{cases} \dfrac{A}{r_{ij}^{8}} - \dfrac{B}{r_{ij}^{6}} & ; \quad 0 < r_{ij} < r_0 \\ 0 & ; \quad r_{ij} \geq r_0 \end{cases} \qquad (3)$$

where B=0.001 [kJmol^{-1}nm^6] and A=0.0022 [kJmol^{-1}nm^8] for CA–CB interactions, and 0.001 for CA–CA ones. r_0 is the "cut–off" value of 0.8 nm.

The residue–specific non–bonded interactions are parameterized as CB–CB contacts. Miyazawa and Jernigan [12] provide a list of pairwise contact energies, obtained on the basis of experimental folded protein structures. Some of these values are positive, some negative, and some equal to 0; since the last case is not desirable for van der Waals type parameterization, we replace it by a value of –0.001 in RT units; this does not lead to significant changes but is more convenient from mathematical point of view. Two cases should be regarded:

I. $\varepsilon_{ij} < 0$. In this "normal" van der Waals type case, the energy is defined as:

$$E_{ij} = \begin{cases} \dfrac{A_{ij}}{r_{ij}^{8}} - \dfrac{B_{ij}}{r_{ij}^{6}} & ; \quad 0 < r_{ij} < r_{ij}^{e} \\ RT\varepsilon_{ij} & ; \quad r_{ij}^{e} \le r_{ij} < r_{ij}^{e} + \Delta r \\ \dfrac{A_{ij}}{(r_{ij} - \Delta r)^{8}} - \dfrac{B_{ij}}{(r_{ij} - \Delta r)^{6}} & ; \quad r_{ij}^{e} + \Delta r \le r_{ij} \le r_{cutoff} \\ 0 & ; \quad r_{ij} > r_{cutoff} \end{cases}$$

where $r^{e}_{ij} = r^{e}_{1/2}(i)+r^{e}_{1/2}(j)$ (i.e. it is the equilibrium or minimum energy distance between pseudo atoms CB_i and CB_j; see Table 1), ε_{ij} denotes contact energy CB_i–CB_j taken from [12], Δr=0.2 nm and r_{cutoff}=50 nm. B_{ij} and A_{ij} are related to ε_{ij} and r^{e}_{ij} by expressions:

$$B_{ij} = -4(r_{ij}^{e})^{6} RT\varepsilon_{ij} \tag{4}$$

$$A_{ij} = 3(r_{ij}^{e})^{2} B_{ij} \tag{5}$$

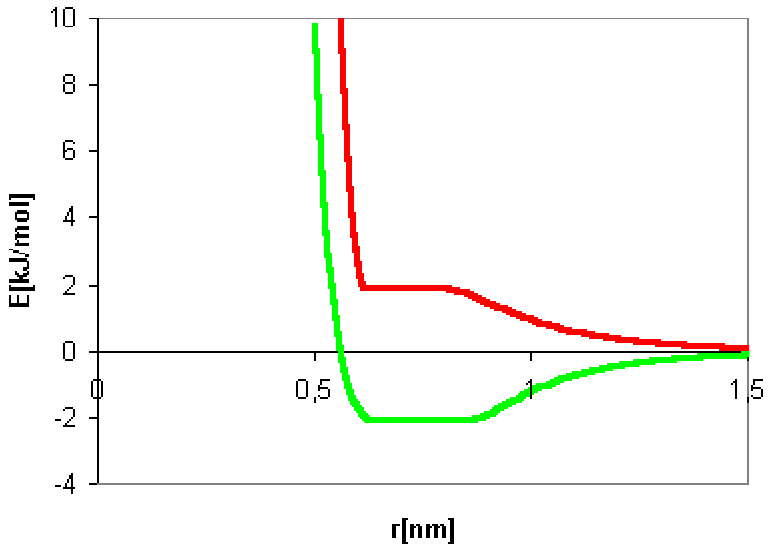


Figure 3. Potential energy function for LEU–LEU (green) and LYS–LYS (red) pairwise contacts.

II. $\varepsilon_{ij} > 0$. The procedure is as follows: primarily, for given values of r^{e}_{ij} and ε_{ij}, A_{ij} and B_{ij} are calculated as in Eqs. (4) and (5), with the negative value $-\varepsilon_{ij}$ instead of ε_{ij}. Then, r^{I}_{ij} is defined as the value of r_{ij}, for which the energy function (3), with the minimum value $-\varepsilon_{ij}$, takes the opposite value of ε_{ij}. Due to the nature of the potential, this is only slightly less than r^{e}_{ij}, obtained analogically as in the case I, for the opposite value of ε_{ij}. Finally, we have:

$$E_{ij} = \begin{cases} \dfrac{A_{ij}}{r_{ij}^{8}} - \dfrac{B_{ij}}{r_{ij}^{6}} & ; \quad 0 < r_{ij} < r_{ij}^{t} \\ RT\,\varepsilon_{ij} & ; \quad r_{ij}^{t} \le r_{ij} < r_{ij}^{t} + \Delta r \\ -\dfrac{A_{ij}}{(r_{ij} - \Delta r)^{8}} + \dfrac{B_{ij}}{(r_{ij} - \Delta r)^{6}} & ; \quad r^{t} + \Delta r \le r_{ij} \le r_{cutoff} \\ 0 & ; \quad r_{ij} > r_{cutoff} \end{cases}$$

where parameters Δr and r_{cutoff} are same as in case I. To illustrate non–bonded energy functions in both cases, examples are presented in Figure 3. All other interaction functions are similar to one of the two types.

The division into segments implies the use of positional restraints for restrained parts of the protein. If the initial coordinates of a given pseudo atom i are x_{0i}, y_{0i} and z_{0i}, and during energy minimization it moves to x_i,y_i,z_i, the restraining energy is given by:

$$E_{restr} = \frac{1}{2} k_{restr} \sum_{i} (x_i - x_{0i})^2 + (y_i - y_{0i})^2 + (z_i - z_{0i})^2 \tag{6}$$

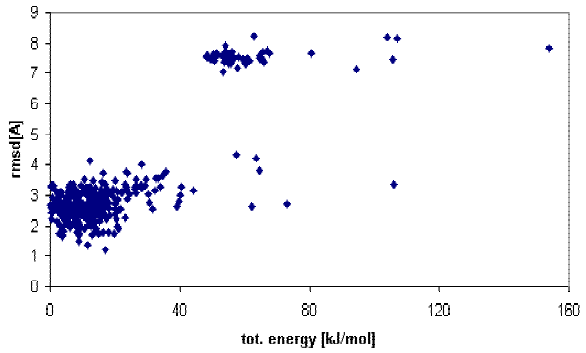
the sum in (6) is over all restrained atoms. For the mobile parts in the reduced representation and with the use of the described force field alone, regular secondary structures, like alpha helices or beta sheets, are only weakly stabilized. In the present test cases, we include information on the secondary structure of the mobile part by employing secondary structure specific distance constraints during energy minimization. That is we assume that it is possible to predict the secondary structure of the mobile segment accurately. If there are M alpha helices 1,2,...M, and the length of each of them is L_1, L_2, ..., L_M, L_i>3, then the constraint energy, E_{constr}, is given by:

$$E_{constr} = \frac{1}{2} \sum_{i=1}^{M} [\sum_{j=1}^{L_i - 2} k_{13} (r_{CA_j CA_{j+2}} - r_{13})^2 + \sum_{j=1}^{L_i - 3} k_{14} (r_{CA_j CA_{j+3}} - r_{14})^2] \tag{7}$$

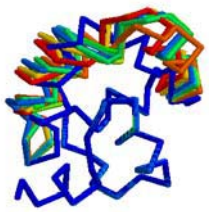
i.e. constraints are imposed on CA pairs of type 1–3 and 1–4. Values of parameters k_{13}, r_{13}, k_{14} and r_{14} from eqs. (6) and (7) are collected in Table 2.

Table 2. Restrained/constrained force field parameters

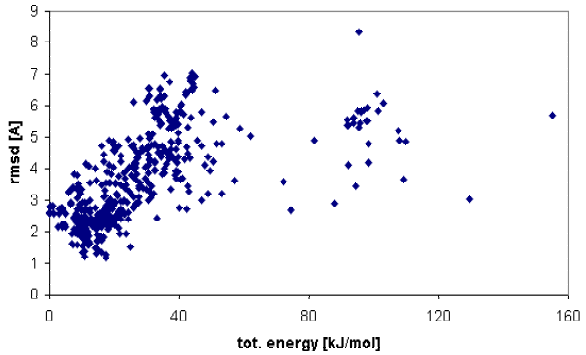
parameter[unit]	value
k_{restr} [kJmol^{-1}nm^{-2}]	2000
k_{13} [kJmol^{-1}nm^{-2}]	5000
r_{13} [nm]	0.545
k_{14} [kJmol^{-1}nm^{-2}]	5000
r_{14} [nm]	0.515



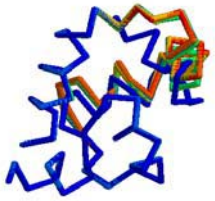
(*a*)



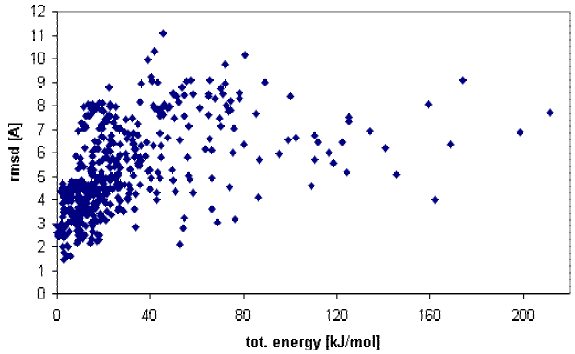
(*e*)



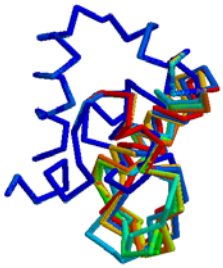
(*b*)



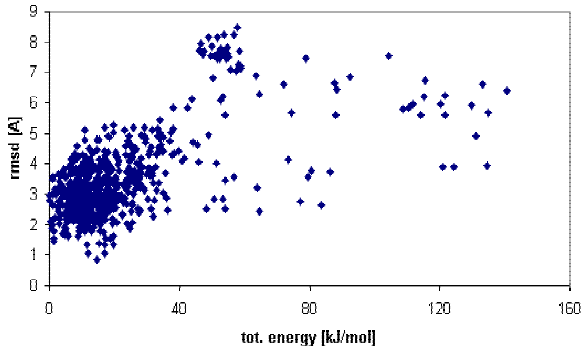
(*f*)



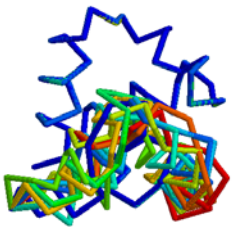
(*c*)



(*g*)



(*d*)



(*h*)

Figure 4. Results for mobile segments I (a, e), II(b,f), III(c,g) and IV(d,h). a–d: Plots of rmsd (CA) with respect to experimental structure vs. total energy e–h: Superposition of 10 structures for the cases I–IV of lowest energy (CA backbones only; restrained blue segments overlap, experimental structure in blue, putative mobile fragments in different colours).

3 RESULTS AND DISCUSSION

The present test protein, 1R69, contains 5 alpha helices (first and last residue in brackets): 1(#2–#12), 2(#17–#24), 3(#28–#35), 4(#45–#51) and 5(#56–#61). We have decided to treat consecutive pairs of alpha helices as mobile segments: 1–2, 2–3, 3–4 and 4–5, including loops between these regular fragments, so that a R–M–R (restraint–mobile–restraint)restraining scheme is in all cases retained (see Methods). The mobile segments for the different test cases are: I(#2–#26), II(#15–#37), III(#26–#53), and IV(#43–#62). The unrestrained (fully mobile) part of the structure varied in length between 20(IV) and 27(III) residues. Note, that this is well beyond the length of loop segments (8–9 residues) that can be accurately modeled based on loop structure databases [2]. For each case, 1000 initial structures were subjected to energy minimization, resulting in 520 energy minimized structures in case I, 647 for II, 532 for III, and 710 for case IV.

Each obtained conformer can be assigned a score (final total energy) and an rmsd (CA atoms only) with respect to the experimental 1R69 structure. Plots of rmsd vs. score are presented in Figure 4a–d. Based on the energy score 10 top scoring structures were pre–selected in each case. These selected structures are presented in Figure 4e–h and in Table 3.

Table 3. Ten structures of lowest energy for each R–M–R case. Total energies are given with respect to the lowest energy minimum in each case.

Mobile segment											
I			II			III			IV		
E [kJ/mol]	rmsd [Å]	# out of 520	E [kJ/mol]	rmsd [Å]	#out of 647	E [kJ/mol]	rmsd [Å]	#out of 532	E [kJ/mol]	rmsd [Å]	#out of 710
0.000	2.47	1	0.000	2.79	7	0.000	2.58	1	0.000	2.80	1
0.113	2.47	1	0.008	2.77	4	0.141	2.87	1	0.017	2.94	1
0.137	2.49	1	0.041	2.78	1	0.289	2.45	1	0.463	2.10	1
0.190	2.60	1	0.048	2.58	7	0.537	2.57	1	0.674	1.87	2
0.289	2.51	1	0.052	2.78	2	0.546	2.58	1	0.736	1.87	1
0.603	2.55	1	0.056	2.58	2	0.553	2.58	1	0.892	2.65	1
0.622	2.55	1	0.065	2.59	2	0.594	2.59	1	0.958	2.99	1
0.664	2.22	1	0.283	2.80	1	0.698	2.50	1	1.091	1.83	1
0.718	3.42	1	0.986	2.79	2	0.711	2.50	1	1.138	3.49	1
0.801	2.51	1	0.996	2.80	3	0.727	2.50	1	1.146	3.55	1

The following factors should be taken into account to assess the method: selectivity (*i.e.*, low energy conformations ought to be close to experiment in rmsd, and high energy ones relatively far), unequivocality (procedure should lead to a limited number of acceptable minima), correlation between energy and rmsd (although, *e.g.*, linear regression is not expected, it would be desirable), and finally, it is desirable to obtain low energy structures close to experiment preferably with an rmsd comparable to the experimental resolution (in case of 1R69: 2Å).

Our results on the present test case reveal a quite good selectivity and reasonable correlation

between energy and rmsd from the experimental structure in all cases. From the results in Table 3 one can estimate an average accuracy of the segment placement of ~2.5–3.0 Å depending on the selected segment. It should be pointed out, that our test protein is relatively small, and the mobile segment contains approximately 1/3 of the whole structure. It is expected that in case of larger structures with less conformational flexibility for the whole structure the prediction for the mobile segment might further improve.

4 CONCLUSIONS

The application of reduced protein models for the prediction of complete proteins structures has so far only been successful in favorable cases still resulting in significant deviations from experiment. In the present study a new concept of a reduced protein modeling approach was introduced that allows rapid generation and selection of putative segment structures for parts of a target protein for which no accurate template structure is available. A key prerequisite of the approach is that the structure of the rest of the protein is either experimentally known or an accurate model structure can be generated using standard homology modeling methods. Focusing the prediction on a protein segment only employing other parts of the protein as conformational constraints to limit the possible mobile segment conformations allows to achieve relatively accurate prediction of protein segment conformations in the present test case (~2.5 Å rmsd from experiment) despite the use of a reduced protein model. The initial tests of the model on a mainly alpha–helical structure are promising. However, further testing of the model on more protein structures and protein classes and use of alternative scoring functions is required to make this approach generally applicable.

Acknowledgment

The financial support of Jena Centre for Bioinformatics (JCB–D2 collaborative project "Homology model based drug design") and a collaboration with Luis Felipe Pineda De Castro (Jenapharm GmbH and Co. KG) are gratefully acknowledged.

5 REFERENCES

[1] M. A. Marti–Renom, A. C. Stuart, A. Fiser, R. Sanchez, F. Melo and A. Sali, Comparative Protein Structure Modeling of Genes and Genomes, *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 291–325.

[2] H. W. van Vlijmen and M. Karplus, PDB–based Protein Loop Prediction: Parameters for Selection and Methods for Optimization, *J. Mol. Biol.* **1997**, *267*, 975–1001.

[3] M. Levitt and A. Warshel, Computer Simulation of Protein Folding, *Nature* **1975**, *253*, 694–698.

[4] J. D. Honeycutt and C. Thirumalai, Metastability of the Folded States of Globular Proteins, *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 3526–3529.

[5] A. Wallqvist and M. Ullner, A Simplified Amino Acid Potential for Use in Structure Predictions of Proteins,

Proteins **1994**, *18*, 267–280.
[6] A. Liwo, S. Oldziej, M. R. Pincus, R. J. Wawak, S. Rackovsky and H. A. Scheraga, A United– Residue Force Field for Off–Latice Protein Structure Simulations, *J. Comp. Chem*. **1997**, *18*, 403–415.
[7] A. Aszodi, R. E. Munro and W. R. Taylor, Protein Modeling by Multiple Sequence Threading and Distance Geometry, *Proteins* **1997**, *Suppl 1*, 38–42.
[8] P. Ulrich, W. Scott, W. F. van Gunsteren and A. E. Torda, Protein Structure Prediction Force Fields: Parametrization with Quasi–Newtonian Dynamics, *Proteins: Struct. Funct. Gen*. **1997**, *27*, 367–384.
[9] A. V. Smith and C. K. Hall, Protein Refolding versus Aggregation: Computer Simulations on an Intermediate–Resolution Protein Model, *J. Mol. Biol*. **2001**, *312*, 187–202.
[10] A. Mondragon, S. Subbiah, S. C. Almo, M. Drottar, and S.C. Harrison, Structure of the Amino– Terminal Domain of Phage 434 Repressor at 2.0 Å Resolution, *J. Mol. Biol*. **1989**, *205*, 189–200.
[11] A. Sali and T. L. Blundell, Comparative Protein Modeling by Satisfaction of Spatial Restraints, *J. Mol. Biol*. **1993**, *234*, 779–815.
[12] S. Miyazawa and R. L. Jernigan, Self– Consistent Estimation of Inter–Residue Protein Contact Energies Based on an Equilibrium Mixture Approximation of Residues, *Proteins: Struct. Funct. Gen*. **1999**, *34*, 49–68.

Biographies

Andrzej Szymoszek, now at UFZ–Center for Environmental Research, Leipzig, Germany, was postdoc at the Institute of Molecular Biotechnology in Jena, Germany. After obtaining a Ph.D. degree in chemistry (2001, University of Wroclaw, Poland, supervisor: Prof. Aleksander Koll), Dr. Andrzej Szymoszek undertook postdoctoral research with Dr. Marjan Vracko at the National Institute of Chemistry in Ljubljana, Slovenia, in the frames of European Union project IMAGETOX (QSAR/QSPR applications in toxicity research). In 2002, Dr. Andrzej Szymoszek moved to Jena, where he started research in frames of the Jena Centre for Bioinformatics project "Homology model based drug design" under scientific supervision of Prof. Martin Zacharias.

Martin Zacharias, is Professor of Computational Biology at the International University Bremen (IUB), Germany. He got his PhD from the Free University Berlin. After postdoctoral periods in the United States and in Berlin he became research group leader at the Institute of Molecular Biotechnology in Jena, Germany, before joining IUB in February 2003.