Available online at www.sciencedirect.com
ScienceDirect
Journal of Steroid Biochemistry & Molecular Biology 103 (2007) 357–360
The Journal of Steroid Biochemistry & Molecular Biology
www.elsevier.com/locate/jsbmb

Steps towards flexible docking: Modeling of three-dimensional structures of the nuclear receptors bound with peptide ligands mimicking co-activators' sequences

Mateusz Kurcinski, Andrzej Kolinski *

Faculty of Chemistry, Warsaw University, Pasteura 1, 02-093 Warsaw, Poland

Received 30 November 2006

Abstract

We developed a fully flexible docking method that uses a reduced lattice representation of protein molecules, adapted for modeling peptide–protein complexes. The CABS model (Carbon Alpha, Carbon Beta, Side Group) employed here, incorporates three pseudo-atoms per residue—Cα, Cβ and the center of the side group instead of full-atomic protein representation. Force field used by CABS was derived from statistical analysis of non-redundant database of protein structures. Application of our method included modeling of the complexes between various nuclear receptors (NRs) and peptide co-activators, for which three-dimensional structures are known. We tried to rebuild the native state of the complexes, starting from separated components. Accuracy of the best obtained models, calculated as coordinate root-mean-square deviation (cRMSD) between the target and the modeled structures, was under 1 Å, which is competitive with experimental methods, such as crystallography or NMR. Forthcoming modeling study should lead to better understanding of mechanisms of macromolecular assembly and will explain co-activators' effects on receptors activity, especially on vitamin D receptor and other nuclear receptors.

Keywords: Flexible docking; Protein interactions; Nuclear receptor–co-activator complex; Drug design

1. Introduction

Nuclear hormone receptors are ligand-activated transcription factors regulating the expression of target genes and thereby affecting cell reproduction, growth and metabolism [1–3]. Presently, it is believed that human genome contains 48 receptors from this family, but only for half of them ligands have been identified [estrogen receptors (ER), androgen receptor (AR), progesterone receptor (PR), glucocorticoid receptor (GR), mineral corticoid receptor (MR), retinoid X receptors (RXR), retinoic acid receptors (RAR), thyroxine hormone receptors (TR), vitamin D receptor (VDR), peroxisome proliferator-activated receptors (PPAR), liver X receptors (LXR), farnesoid receptor (FXR) and steroid xenobiotic receptor (SXR)]. From the structural point of view all nuclear receptors form a superfamily (according to SCOP [4]) with highly conserved topology of structural motives despite that they bind different ligands. NRs consist of an N-terminal region responsible for ligand-independent transcriptional activation (AF-1), DNA-binding domain (DBD) containing motif of two zinc fingers and C-terminus including ligand-specific binding domain (LBD) and flexible hinge which "locks" ligand upon binding. Usually nuclear receptors are investigated in respect of their interaction only with primary ligands [5–7], but they also form complexes with other molecules. In *holo*-form NRs bind with various cell-specific co-activators, which link receptor with the RNA polymerase II—a gear in transcriptional mechanism, while in *apo*-form NRs form complexes with co-repressors and act as transcriptional suppressors. Abundance of functions of different nuclear receptors and structural similarity between them at the same time makes the NR class very promising pharmacological target.

* Corresponding author. Tel.: +48 22 822 02 11x320; fax: +48 22 822 59 96.
E-mail address: kolinski@chem.uw.edu.pl (A. Kolinski).

0960-0760/$ – see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsbmb.2006.12.059

2. Materials and methods

In present work, we describe an application of a recently developed fully flexible docking algorithm to a set of protein complexes from the NR superfamily for which three-dimensional structures are known (so-called 'bound' docking). Our method incorporates the reduced CABS model which has been initially designed for single chain protein folding and performed well in many applications [8]. It has been previously described in great details [9], here we only mention its most basic features.

2.1. Model description

Model CABS is a high resolution reduced model. It assumes protein representation as a three interaction centers per residue (Cα, Cβ, side group). Cα atoms are located on the simple cubic lattice with lattice unit equal to 0.61 Å. Cβ and side group atoms are located off the lattice and their positions are defined by three consecutive Cα atoms. CABS force field was constructed by statistical analysis of non-redundant database of experimentally solved protein structures in the form of histograms reflecting ensembles of various structural properties. It covers most typical interactions such as hydrogen bonds, electrostatics, hydrophobic attraction, but also protein-specific ones: disulfide bridges, centrosymmetrical potential (reflecting the hydrophobic effect), or biases towards regular secondary structure. Sampling of the conformational space is controlled by the Replica Exchange Monte Carlo scheme [10]. Molecules undergo small random conformational changes which are accepted according to the Metropolis Criterion [11]. Additionally, simulation runs in several replicas (copies) in different temperatures and every given amount of simulation steps coordinates of the replicas are exchanged with probability proportional $\exp\{-\Delta E \Delta\beta\}$.

2.2. Protein complexes

We selected from the PDB database 10 high-resolution structures of various nuclear receptor complexes to study the quality of our docking procedure. Starting point for the simulations assumed that both molecules (receptor and co-activator) in their native states were shifted apart from each other to an arbitrary distance of 40 Å between their centers of gravity. Ligand molecules were not explicitly present in simulations, but their influence on the system was incorporated in structural restraints imposed on receptor molecules. None restraints were imposed on co-activators' molecules.

2.3. Clustering and scoring

Trajectories obtained during simulations contained couple of thousands frames, from which only few could be selected as final structures. Since our algorithm uses random search in conformational space, the last frame is not usually the best one. Moreover, the frame with the lowest energy also cannot be simply chosen as the most accurate model, because our method uses statistical force field, where energy does not correlate straightforwardly with the real free energy. In order to select the final structures hierarchical clustering was applied to every trajectory. We used the HCPM [12] program, with the single-link clustering procedure and coordinate root-mean-square deviation (cRMSD) as a similarity measure. Structure of the best representative of the cluster was selected as a frame closest (in the means of cRMSD) to the centroid of the cluster.

Except for the minimal cRMSD, calculated for the whole complex, the quality of the predicted complex structures was assessed in terms of the following measures [13]:

Ligand cRMSD, defined as the cRMSD calculated only for the ligand backbone after superimposition of the receptor structures.

Fraction of native contacts, defined as the number of correct contacts between receptor's and ligand's residues in the predicted complex, divided by the number of contacts in the native structure of the complex. A pair of residues is arbitrary considered to be in contact if distance between their Cα atoms is less then 10 Å.

3. Results and discussion

Final models were selected from trajectories in two different manners. At first every frame of the trajectory was compared with the native structure from the PDB database in

Table 1
Best frames selected from the trajectories compared with crystallographic structure

PDB code	Chains	NR class	Minimal cRMSD (Å)	Ligand cRMSD (Å)	Fraction of native contacts
1KKQ	A E	PPARα	4.26	14.88	0.11
1KV6	A C	ERR	1.48	1.71	0.68
1M2Z	A B	GR	2.99	8.46	0.18
1MVC	A B	RXRα	0.49	1.20	0.94
1NQ7	A B	RORβ	0.42	0.86	0.90
1NRL	A C	PXR	5.09	22.64	0.00
1OSV	B D	FXR	1.37	3.94	0.36
1RJK	A C	VDR	0.58	1.47	0.63
1XB7	A P	ERRα	7.33	19.09	0.02
3ERD	A C	ERα	0.87	2.03	0.64

Table 2
Final models selected in clustering procedure, compared with crystallographic structures

PDB code	Cluster	Minimal cRMSD (Å)	Ligand cRMSD (Å)	Fraction of native contacts
1KKQ	1	5.76	15.20	0.09
	2	8.71	29.67	0.00
	3	**4.99**	**17.14**	**0.15**
1KV6	**1**	**2.29**	**4.24**	**0.58**
	2	8.34	40.33	0.00
	3	7.40	35.50	0.00
1M2Z	1	10.60	39.30	0.00
	2	9.82	36.37	0.00
	3	**9.08**	**33.87**	**0.00**
1MVC	1	0.95	3.86	0.70
	2	0.83	3.19	0.62
	3	**0.74**	**2.78**	**0.74**
1NQ7	**1**	**0.47**	**0.89**	**0.90**
	2	0.70	2.84	0.74
	3	3.71	19.45	0.06
1NRL	**1**	**5.97**	**26.41**	**0.00**
	2	6.24	27.59	0.00
	3	6.43	28.24	0.00
1OSV	**1**	**1.70**	**4.96**	**0.31**
	2	4.91	21.93	0.00
	3	5.76	26.37	0.00
1RJK	**1**	**0.78**	**3.08**	**0.72**
	2	2.33	10.93	0.30
	3	2.07	9.60	0.32
1XB7	1	10.46	40.28	0.00
	2	8.08	25.50	0.00
	3	**7.96**	**24.37**	**0.00**
3ERD	1	6.61	33.07	0.00
	2	**1.72**	**7.41**	**0.23**
	3	6.99	34.98	0.00

Best models are presented in bold.

the means described in the previous paragraph and the final model was chosen as the frame with the lowest cRMSD to the native structure. Results are presented in Table 1. Afterwards we applied different method of selecting final models from the trajectory, referring to the cases when native structure of the complex is unknown and selection of the final model must be made using non-comparative techniques. Clustering procedure was applied to trajectory and final models were chosen as representative structures from the three biggest clusters (containing the largest number of frames). Quality of obtained models was assessed by comparing them with the native structure. For five models the top-scored cluster was at the same time the biggest one. In remaining cases it was among the three biggest. Results are presented in Table 2.

In three cases (1MVC, 1NQ7, 1RJK) obtained models may be considered as very good (cRMSDs < 1 Å), in other three (1KV6, 1OSV, 3ERD) as good (cRMSD < 2.5 Å) (Figs. 1 and 2). The remaining four models are inaccurate (cRMSD ≈ 5–10 Å), however two of them (1KKQ, 1NRL) still may be a source of qualitative structural information about the location of the binding site, since in these cases co-activators were docked in correct site on the surface of the receptor, but were wrongly oriented. It is worth mentioning that no additional information about location of the binding site was used in modeling. It indicates that our algorithm is able to utilize structural information contained in receptor's

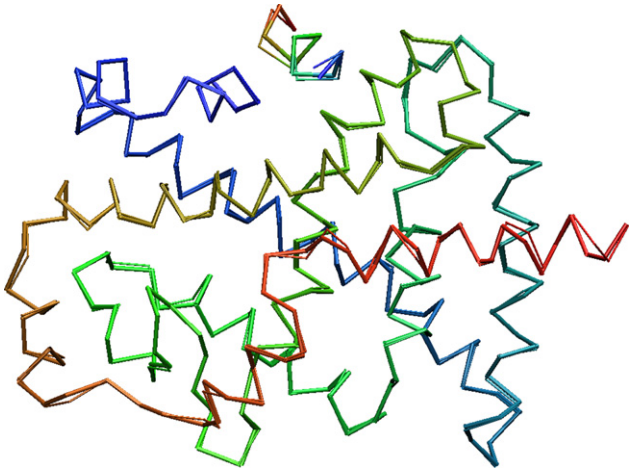

Fig. 1. Best model (thick lines) of 1NQ7 in the alpha carbon representation superimposed onto crystallographic structure (thin lines).

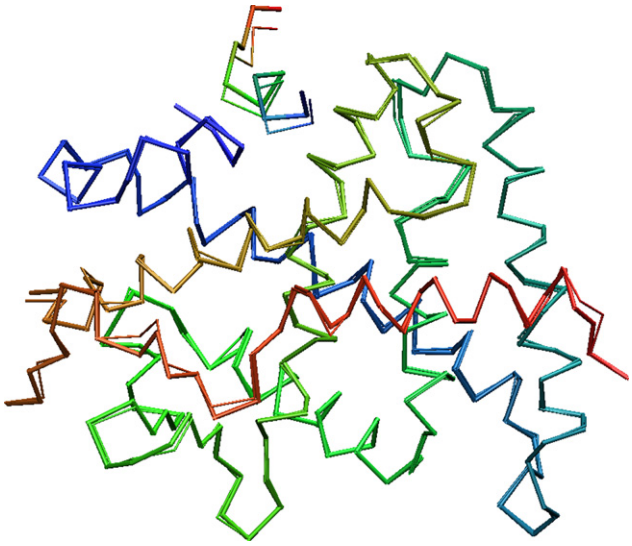

Fig. 2. Best model (thick lines) of 1RJK in the alpha carbon representation superimposed onto crystallographic structure (thin lines).

structure without calculating any sort of molecular surface, but only by efficient exploration of the energy landscape.

We discovered that for those complexes that were modeled with the highest accuracy the biggest cluster contained almost half of the trajectory frames, while for those worse modeled only about 20% frames belonged to the first cluster. This may be a strong premise to verify if the model is correct in 'unbound' docking cases, where no native structure to compare with is available. Here we present application of our algorithm only to 10 examples of protein complexes, which are additionally close homologues. In order to verify above mentioned thesis, wider experiment must be conducted, which would cover at least couple of tens various proteins.

4. Conclusion

We developed a new algorithm for fully flexible docking of peptides and proteins. It is based on previously described CABS mesoscopic modeling tool which was successfully used for study of protein dynamics and thermodynamics [14,15] and prediction of protein structures [16]. It has been shown here that the new modeling tool is already capable of producing correct high-resolution structures of protein–peptide complexes. This should be very important for understanding of protein interactions, signaling pathways and computed aided design of new drugs. Work in progress aims on designing of a complete protein interactions modeling tool, which could be used to simulate small ligand docking, macromolecular assembly of protein domains and interactions between proteins and nucleic acids.

Acknowledgment

Computational part of this work was done using the computer cluster at the Computing Center of Faculty of Chemistry, Warsaw University.

References

[1] R.M. Evans, The steroid and thyroid hormone receptor superfamily, Science 240 (4854) (1988) 889–895.
[2] D.J. Mangelsdorf, et al., The nuclear receptor superfamily: the second decade, Cell 83 (6) (1995) 835–839.
[3] A. Chawla, et al., Nuclear receptors and lipid physiology: opening the X-files, Science 294 (5548) (2001) 1866–1870.
[4] A.G. Murzin, et al., SCOP: a structural classification of proteins database for the investigation of sequences and structures, J. Mol. Biol. 247 (4) (1995) 536–540.
[5] P. Rotkiewicz, et al., Model of three-dimensional structure of vitamin D receptor and its binding mechanism with 1alpha,25-dihydroxyvitamin D(3), Proteins 44 (3) (2001) 188–199.
[6] R.R. Sicinski, et al., 2-Ethyl and 2-ethylidene analogues of 1alpha,25-dihydroxy-19-norvitamin D(3): synthesis, conformational analysis, biological activities, and docking to the modeled rVDR ligand binding domain, J. Med. Chem. 45 (16) (2002) 3366–3380.
[7] W. Sicinska, P. Rotkiewicz, H.F. DeLuca, Model of three-dimensional structure of VDR bound with Vitamin D3 analogs substituted at carbon-2, J. Steroid Biochem. Mol. Biol. 89–90 (1–5) (2004) 107–110.
[8] A. Kolinski, J.M. Bujnicki, Generalized protein structure prediction based on combination of fold-recognition with de novo folding and evaluation of models, Proteins 61 (Suppl. 7) (2005) 84–90.
[9] A. Kolinski, Protein modeling and structure prediction with a reduced representation, Acta Biochim. Pol. 51 (2) (2004) 349–371.
[10] R.H. Swendsen, J.S. Wang, Replica Monte Carlo simulation of spin glasses, Phys. Rev. Lett. 57 (21) (1986) 2607–2609.
[11] N. Metropolis, et al., Equation of state calculations by fast computing machines, J. Chem. Phys. 51 (1953) 1087–1092.
[12] D. Gront, A. Kolinski, HCPM–program for hierarchical clustering of protein models, Bioinformatics 21 (14) (2005) 3179–3180.
[13] S. Vajda, C.J. Camacho, Protein-protein docking: is the glass half-full or half-empty? Trends Biotechnol. 22 (3) (2004) 110–116.
[14] D. Ekonomiuk, M. Kielbasinski, A. Kolinski, Protein modeling with reduced representation: statistical potentials and protein folding mechanism, Acta Biochim. Pol. 52 (4) (2005) 741–748.
[15] S. Kmiecik, et al., Denatured proteins and early folding intermediates simulated in a reduced conformational space, Acta Biochim. Pol. 53 (1) (2006) 131–144.
[16] M. Boniecki, et al., Protein fragment reconstruction using various modeling techniques, J. Comput. Aided Mol. Des. 17 (11) (2003) 725–738.