Published in final edited form as:

Bioorg Med Chem Lett. 2011 October 15; 21(20): 6081–6084. doi:10.1016/j.bmcl.2011.08.051.

Beyond the frog: The evolution of homology models of human IKK β

Nicholas Y. Palermo^a and Amarnath Natarajan^{a,*}

^aEppley Institute for Cancer Research, University of Nebraska Medical Center, Omaha, NE, 68198, USA

Abstract

Nuclear Factor κ B is implicated in tumor progression and chronic inflammatory diseases and is regulated by I κ B kinase β (IKK β). The crystal structure of IKK β has been recently solved for *xenopus laevis*. Homology models of human IKK β have been developed prior to and after the crystal structure was solved. Here, we compare four models of human IKK β and evaluate their performance in both broad and focused library docking studies.

Keywords

I κ B kinase β ; Homology modeling; NF- κ B; AutoDock Vina; YASARA

Constitutively active Nuclear Factor κ B (NF- κ B) has been implicated in tumor progression as well as multiple chronic inflammatory diseases.¹ NF- κ B is inactive when it is in complex with its inhibitor, I κ B.² I κ B kinase β (IKK β) is a serine-threonine kinase^{3–4} which activates the transcription factor NF- κ B by phosphorylating I κ B. The phosphorylated I κ B is ubiquitinated and proteasomally degraded.⁵ This leads to the dissociation of the I κ B/NF- κ B complex, allowing NF- κ B to translocate to the nucleus.⁶ Once NF- κ B is in the nucleus, it modulates the transcription of genes implicated in proliferation, angiogenesis and chemoresistance.^{6,7} Consequently, IKK β is an important therapeutic target in the NF- κ B pathway and inhibitors of IKK β are being actively developed by drug companies including Bayer^{8–10}, Amgen^{11,12}, Pfizer¹³ and Daiichi Sankyo.¹⁴

In silico structure-based drug design methodologies could provide a significant advantage to discovering novel compounds which target IKK β , but no experimentally determined structure of human IKK β is reported. However, the crystal structure of *xenopus laevis* IKK β , which has 74.0% sequence similarity with human IKK β , has been solved recently.¹⁵ This structure reveals that the first 666 residues of IKK β comprise a kinase domain, ubiquitin-like domain and an α -helical scaffold/dimerization domain.

Homology modeling is one option for obtaining 3D structures of proteins that have not yet been solved. 3D structures of druggable targets can be used for virtual screening studies and lead compound discovery. In recent years, homology modeling methodologies have improved and the technique is becoming more widely used for drug discovery and structural

© 2011 Elsevier Ltd. All rights reserved.

*Corresponding author. Tel.: +1-402-559-3793; fax: +1-402-559-8270; anatarajan@unmc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

studies.^{16–17} Uses for homology models include but are not limited to structure-based virtual screening for lead compounds, optimization of compounds, assessing the biological role of mutations and post translational modifications.¹⁶ In the present study, we generated homology models of IKK β before and after the publication of the *xenopus* IKK β structure and compared their utility for hit identification and hit-to-lead optimization.

Four homology models were investigated, two from the SWISS-MODEL^{18–20} repository and two generated in our laboratory (Table 1). Models M1 and M2 were generated before the *xenopus* IKK β structure was published and models M3 and M4 use the published structure as a template.

The YASARA Structure²¹ program package was used to generate homology models M2 and M4. YASARA was chosen as the homology modeling package because it has performed well in Critical Assessment of Techniques for Protein Structure Prediction 8 (CASP8) and uses unique knowledge-based potential force fields to better predict molecular structure.²² The IKK β FASTA sequence, accession number O14920, was obtained from Uniprot.²³ Template structures were selected on the basis of having the lowest E-value after six iterations of the PSI-BLAST program. Given that the PSI-BLAST runs yielded 654 possible templates with an E-value less than 0.1, seven structures, representing the top 1%, were chosen as templates. Each template structure was required to have a unique sequence. Up to five different variants were allowed per template if the alignment of the template sequence on the target sequence was ambiguous. The oligomerization state parameter was set to 2 to account for potential dimeric structures. The final model was selected on the basis of having the best Z-score after a combined steepest descent and simulated annealing minimization. Each model was then further refined using the md_refine macro included with YASARA. The YASARA2 force field was chosen for the refinement simulations²².

The backbone structures of the crystal structure and the four homology models generated are represented in Figure 1. All of the homology models include the full-length kinase domain, residues 16–307, with the exception of M1, which ends at residue 304. Only models M3 and M4 accurately represent the ubiquitin-like and α -helical scaffold/dimerization domains, albeit M4 has a noticeable kink beginning at residue Asn 470 and ending at residue Asn 528, which is due to the kinked conformation being more energetically favorable than the extended conformation, in the context of an H₂O-filled simulation environment.

The quality Z-score of the Dihedral, 1D and 3D packing components of each model are in Table 2, with higher numbers indicating better model quality. The overall Z-score is computed as the weighted average of $0.145 \cdot \text{Dihedral} + 0.390 \cdot \text{Packing1D} + 0.465 \cdot \text{Packing3D}$. The YASARA-generated models, M2 and M4, show substantially increased quality compared to those generated by SWISS-MODEL.

The RMSD of the residues of the kinase domain, 1–307, with respect to the *xenopus* model, are less in models M1 and M3, 1.86 Å and 1.10 Å, respectively, than that of models M2 and M4, 1.99 Å and 1.75 Å, respectively. The position of the residues of the activation loop of the kinase domain, residues 166–192, deviate from the crystal substantially in M1 and M2 (Figure 2).

The RMSD of the activation loop residues are 6.67 Å, 10.83 Å, 2.13 Å, and 2.25 Å in models M1, M2, M3 and M4, respectively. The positions of the side-chain in the ATP binding site of IKK β vary considerably between the models (Figure 3). Models M1, M2 and M3 all have side-chains which are positioned in the binding site, whereas M4 most closely resembles the crystal structure. In model M1, residue Tyr98 has been displaced approximately 1.4 Å and the χ^1 dihedral angle of Met96 has been rotated -45.7 degrees, so that both side-chains are in the binding pocket. Residue Cys99 of M2 has been displaced

approximately 1.5 Å into the binding pocket. Residues Cys99 and Ile165 of M3 have been displaced 0.90 Å and rotated -39.6 degrees, respectively, so that both side-chains are in the binding pocket. Model M4 has no significant changes to side-chain position or orientation which affect the binding site, because the homology modeling package implemented in YASARA includes any ligands found in the template structure during the modeling process, which yields a better representation of the active conformation of the binding site.

A two-pronged approach was chosen to explore the utility of the models for drug discovery. The ability of the models to identify hits was tested by virtually screening the Prestwick chemical library, which was seeded with experimentally verified IKK β inhibitors. Success was assessed on the basis of the number of seeded compounds which were in the top 10% of the virtual screen. The ability of the models to be used for hit-to-lead optimization was tested by calculating the energy of binding of individual inhibitors from reported sets of focused libraries and correlating experimentally and calculated affinities.

AutoDock Vina was used for the broad and focused library docking calculations.²⁴ All four models were used for both libraries. The minimum size of each docking box was 20 Å $\times$ 20 Å $\times$ 20 Å and exhaustiveness was set to 16. The receptor was treated as rigid.

The Prestwick Chemical Library, a 1200 member library of known drugs, was used for hit identification. Seven compounds from the Bayer and Amgen libraries, which had the highest biological activity, were seeded into the Prestwick chemical library. Compounds with more than 9 torsional angles, the highest number of torsional angles in the seeded compounds, were not docked; the total number of screened ligands was 992. Effectiveness of each model was determined by counting the number of seeded compounds which ranked in the top 10% best calculated ΔG of binding. The results of the hit identification screen are reported in Table 3. M4 is the only model in which more than half of the seeded compounds were successfully counted as hits, making it the model of choice for hit identification.

The focused library used for the hit-to-lead optimization validation was composed of IKK β inhibitors developed by Bayer and Amgen, 132 compounds in all. The compounds in each publication were considered a separate data set (Table 4). Compounds were chosen for docking only if their experimentally determined IC₅₀ was within measurable bounds. The effectiveness of each model was determined by calculating the correlation between calculated ΔG of binding and experimentally determined IC₅₀. Correlation values were calculated using the Gnumeric program (<http://www.gnumeric.org>).

The correlation (Table 5) between experimentally determined IC₅₀ and calculated ΔG (Table S1) was generally below 0.5. Models M2 and M4 appear to be the most accurate models because their correlation values are consistently positive for all data sets. The low R-values overall are probably due to a lack of adequate energy terms in the Vina scoring function. The calculated ΔG values of all data points were found to fall within the ± 2.85 kcal mol⁻¹ of the median data point, which is the standard error of Vina. This study suggests at least for our system Vina has limited use from a ligand optimization perspective. A more general reason for low correlation that is endemic to virtual screening is that the receptor can be flexible and therefore higher accuracy could be obtained by modeling it as such, albeit at increased computational cost.

In summary, this investigation has revealed that the quality of human IKK β homology models has increased substantially since the publication of the *xenopus* model. Models M1 and M2, which predate the crystal structure, do not have a correctly positioned kinase domain activation loop, which is likely to be a critical factor in activity studies. Models M3 and M4 more accurately reproduce the backbone structure of the crystal, but only M4 replicates the conformation of the binding site. Therefore, model M4, generated by

YASARA, is a suitable model for broad screening. Our studies also suggest that the model is more useful for the identification of hits as opposed to the hit-to-lead optimization.

Experimental

The full details of the homology modeling experiments of models M2 and M4 can be accessed in the supplementary material. Note that the reported Z-scores are before the model is optimized using molecular dynamics.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH T32CA009476 and in part by the Eppley Cancer Center Pilot Grant. We would like to thank Dr. Vashti C. Bryant for critically reviewing the manuscript.

References

1. Hayden M, Ghosh S. Cell. 2008; 132:344. [PubMed: 18267068]
2. Baeuerle P, Baltimore D. Science. 1988; 28:540. [PubMed: 3140380]
3. Scheidereit C. Oncogene. 2006; 25:6685. [PubMed: 17072322]
4. Vallabhapurapu S, Karin M. Annu Rev Immunol. 2009; 27:693. [PubMed: 19302050]
5. Mercurio F, Zhu H, Murray B, Shevchenko A, Bennett B, Li J, Young D, Barbosa M, Mann M, Manning A, Rao A. Science. 1997; 278:860. [PubMed: 9346484]
6. Woronicz J, Gao X, Cao Z, Rothe M, Goeddel D. Science. 1997; 278:866. [PubMed: 9346485]
7. Cui J, Zhu L, Xia X, Wang H, Legras X, Hong J, Ji J, Shen P, Zheng S, Chen Z, Wang F. Cell. 2010; 141:483. [PubMed: 20434986]
8. Murata T, Shimada M, Sakakibara S, Yoshino T, Kadono H, Masuda T, Shimazaki M, Shintani T, Fuchikami K, Sakai K, Inbe H, Takeshita K, Niki T, Umeda M, Bacon K, Ziegelbauer K, Lowinger T. Bioorg Med Chem Lett. 2003; 13:913. [PubMed: 12617920]
9. Murata T, Shimada M, Kadono H, Sakakibara S, Yoshino T, Masuda T, Shimazaki M, Shintani T, Fuchikami K, Bacon K, Ziegelbauer K, Lowinger T. Bioorg Med Chem Lett. 2004; 14:4013. [PubMed: 15225717]
10. Murata T, Shimada M, Sakakibara S, Yoshino T, Masuda T, Shintani T, Sato H, Koriyama Y, Fukushima K, Nunami N, Yamauchi M, Fuchikami K, Komura H, Watanabe A, Ziegelbauer K, Bacon K, Lowinger T. Bioorg Med Chem Lett. 2004; 14:4019. [PubMed: 15225718]
11. Cushing T, Baichal V, Berry K, Billedeau R, Bordunov V, Broka C, Cardozo M, Cheng P, Clark D, Dalrympie S, DeGraffenreid M, Gill A, Hao X, Hawley R, He X, Jaen J, Labadie S, Labelle M, Lehel C, Lu PP, McIntosh J, Miao S, Parast C, Shin Y, Sjogren E, Smith ML, Talamas F, Tonn G, Walker K, Walker N, Wesche H, Whitehead C, Wright M, Browner M. Bioorg Med Chem Lett. 2011; 21:417. [PubMed: 21074993]
12. Cushing T, Baichal V, Berry K, Billedeau R, Bordunov V, Broka C, Browner M, Cardozo M, Cheng P, Clark D, Dalrympie S, DeGraffenreid M, Gill A, Hao X, Hawley R, He X, Labadie S, Labelle M, Lehel C, Lu P-P, McIntosh J, Miao S, Parast C, Shin Y, Sjogren E, Smith M-L, Talamas F, Tonn G, Walker K, Walker N, Wesche H, Whitehead C, Wright M, Jaen J. Bioorg Med Chem Lett. 2011; 21:423. [PubMed: 21074992]
13. Crombie A, Sum FW, Powell D, Hopper D, Torres N, Berger D, Zhang Y, Gavriil M, Sadler T, Arndt K. Bioorg Med Chem Lett. 2010; 20:3821.
14. Shimizu H, Tanaka S, Toki T, Yasumatsu I, Akimoto T, Morishita K, Yamasaki T, Yasukochi T, Iimura S. Bioorg Med Chem Lett. 2010; 20:5113. [PubMed: 20675134]
15. Xu G, Lo Y, Li Q, Napolitano G, Wu X, Jiang X, Dreano M, Karin M, Wu H. Nature. 2011; 472:325. [PubMed: 21423167]

16. Cavasotto C, Phatak S. Drug Discovery Today. 2009; 14:676. [PubMed: 19422931]
17. Venselaar H, Joosten R, Vroiling B, Baakman C, Hekkelman M, Krieger E, Vriend G. Eur Biophys J. 2010; 39:551. [PubMed: 19718498]
18. Arnold K, Bordoli L, Kopp J, Schwede T. Bioinformatics. 2006; 22:195. [PubMed: 16301204]
19. Kiefer F, Arnold K, Kunzli M, Bordoli L, Schwede T. Nucleic Acids Research. 2009; 37:D387. [PubMed: 18931379]
20. Peitsch MC. Nature Biotechnology. 1995:658.
21. YASARA. [Http://www.yasara.org](http://www.yasara.org)
22. Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson J, Tyka M, Baker D, Karplus K. Proteins. 2009; 77(Suppl 9):114. [PubMed: 19768677]
23. The UniProt Consortium. Nucleic Acids Research. 2011; 39:D214. [PubMed: 21051339]
24. Trott O, Olson AJ. Comp Chem. 2010; 31:455.

Figure 1.
Cartoon representation of the backbone of the crystal structure and the homology models of IKK β .

Figure 2.
Tube representation of the backbone of the activation loop residues of the kinase domain.
The crystal structure is in green. M1, M2, M3 and M4 are in red, magenta, yellow and cyan,
respectively.

Figure 3. Differences in side-chain positions of the models of IKK β . The gray surface representation is of the ATP binding site of the *xenopus* crystal structure. M1, M2 and M3 are in red, green and yellow, respectively.

Table 1Homology models of IKK β .

Model	Source	Source PDB	Residues
M0	X-ray structure	-	16-666
M1	SWISS-MODEL	1cm8	12-304
M2	YASARA	3ku2	1-554
M3	SWISS-MODEL	3qa8	16-666
M4	YASARA	3qad	6-647

Table 2Z-scores of homology models of IKK β .

Model	Dihedral	Packing1D	Packing3D	Overall
M1	-2.39	-4.39	-4.52	-4.16
M2	0.92	-2.30	-1.68	-1.54
M3	-7.44	-6.57	-7.25	-7.01
M4	0.63	-2.71	-1.77	-1.78

Table 3

Number of seeded compound hits in the top 10% of the hit identification screen.

Model	Seeded in top 10%
M1	1
M2	2
M3	1
M4	4

Table 4

Data sets for focused library molecular docking calculations.

Data Set	Number of compounds	Developer	Reference
Set1	14	Bayer	13
Set2	17	Bayer	14
Set3	29	Bayer	15
Set4	40	Amgen	16
Set5	32	Amgen	17

Table 5

Correlation of ΔG and IC_{50} .

Model	R_{set1}	R_{set2}	R_{set3}	R_{set4}	R_{set5}
M1	-0.11	0.09	-0.3	0.17	0.37
M2	0.34	0.18	0.67	0.25	0.19
M3	0.34	0.16	-0.30	-0.31	-0.13
M4	0.36	0.25	0.32	0.14	0.19