Quantitative modeling and analysis of the transforming growth factor β signaling pathway

Seung-Wook Chung, Fayth L. Miles, Carlton R. Cooper, Mary C. Farach-Carson, and Babatunde A. Ogunnaike
University of Delaware, Newark, DE, USA

Introduction

Transforming growth factor-β (TGF-β) proteins are members of a superfamily of secreted cytokines that control a diverse array of cellular processes including cell proliferation, differentiation, motility, adhesion, angiogenesis, apoptosis, and immune surveillance. The TGF-β signaling cascade begins when extracellular TGF-β binds to and brings together Type I and Type II TGF-β receptor serine/threonine kinases on the cell surface, whereby the Type II receptor phosphorylates and activates the Type I receptor. The activated Type I receptor, in turn, propagates the signal through phosphorylation of receptor-bound (R-)Smad transcription factors (Smad2/3 and Smad1/5/8). The activated R-Smads form hetero-oligomers with Smad4 and rapidly translocate into the nucleus, undergoing continuous nucleocytoplasmic shuttling by interacting with the nuclear pore complex. Once in the nucleus, the activated Smad complexes bind to specific promoters and ultimately regulate expression of target genes, generating approximately five hundred gene responses in a cell- and context-specific manner[1, 2].

The TGF-β signaling pathway has become an attractive but problematic target for oncology drug development because of its apparently paradoxical roles in tumorigenesis and metastasis. In normal and early phase tumorigenic epithelial cells, TGF-β functions as a potent tumor suppressor primarily by inducing cell cycle arrest and apoptosis. However, in the intermediate and late stages of carcinogenesis, tumor cells become resistant to the growth inhibitory effects of TGF-β and show elevated expression of TGF-β. The ligand is over-expressed in clinical cancer samples, with increasing levels correlating with poor clinical outcomes. The role of TGF-β therefore *appears* to become one of tumor promotion, apparently supporting growth, subverting the immune system, and also facilitating, invasion, epithelial to mesenchymal transition (EMT), and angiogenesis. This finding has created the widely held perception that TGF-β acts as a tumor *promoter* in advanced tumorigenesis and metastasis[3]. Although it is known that most cancer cell lines representing the entire spectrum of tumor progression have active TGF-β signaling pathways, detailed mechanisms of how a single stimulus, TGF-β, induces such a diverse array of responses during cancer progression are not known. This is primarily due to the complexity of the signaling cascade system in which a variety of signaling components changing dynamically over different time scales interact with one another. Since quantitative understanding and analysis of such a complex regulatory circuit are not possible via qualitative human intuition alone, mathematical descriptions that lead to predictive models are necessary.

To improve our understanding of complex TGF-β signaling quantitatively, we have developed a comprehensive, dynamic model of the canonical TGF-β pathway via Smad transcription factors, based on the most up-to-date information available in the literature. Through simulation and model analysis, our model provides insight into the signal-response relationship between the binding of TGF-β to its receptor at the cell surface and the activation of downstream effectors in the signaling cascade. In particular, we use the model to carry out

"*in-silico* mutations" from which we generate several hypotheses regarding potential mechanisms for how TGF-β's tumor-suppressive roles may appear to morph into tumor-promoting roles.

Model Formulation

Our model incorporates the following essential molecular processes: (i) sequential activation of receptors induced by TGF-β; (ii) receptor internalization to endosomes and recycling; (iii) R-Smad phosphorylation by receptor complex; (iv) R-Smad-Smad4 complex formation in the cytoplasm and nucleus; (v) nucleocytoplasmic shuttling of Smads; (vi) dissociation and dephosphorylation of activated Smads; (vii) constitutive and ligand-induced degradation of proteins; and (viii) protein synthesis.

The components of the overall TGF-β signaling pathway as featured in our model are depicted in Figure 1. The resulting model is a system of 17 non-linear ordinary differential equations (ODEs) with 37 kinetic parameters arising from chemical reactions represented by mass action kinetics (Table 1).

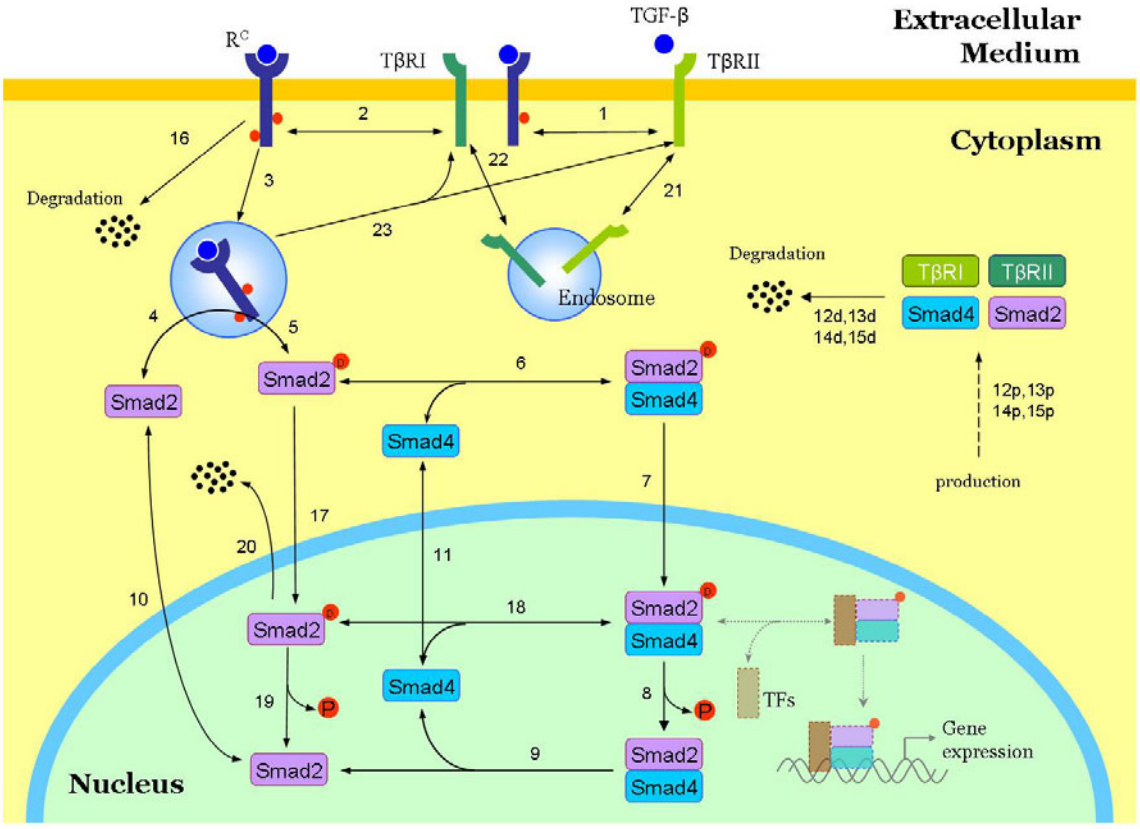

Figure 1. Schematic representation of the pathway components in the model.

Table 1. Model Equations

$$v_1 = k_{1a}[TGF\beta][RII] - k_{1d}[TGF\beta : RII] \qquad v_2 = k_{2a}[TGF\beta : RII][RI] - k_{2d}[R^C]$$

$$v_3 = k_{3\text{int}}[R^C] \qquad v_4 = k_{4a}[R_{in}^C][S2_{cyt}] - k_{4d}[R_{in}^C : S2_{cyt}]$$

$$v_5 = k_{5cat}[R_{in}^C : S2_{cyt}] \qquad v_6 = k_{6a}[pS2_{cyt}][S4_{cyt}] - k_{6d}[pS2S4_{cyt}]$$

$$v_7 = k_{7imp}[pS2S4_{cyt}] \qquad v_8 = k_{8dp}[pS2S4_{nuc}]$$

$$v_9 = k_{9d}[S2S4_{nuc}] \qquad v_{10} = k_{10imp}[S2_{cyt}] - k_{10\text{exp}}[S2_{nuc}]$$

$$v_{11} = k_{11imp}[S4_{cyt}] - k_{11\text{exp}}[S4_{nuc}] \qquad v_{12} = k_{12syn} - k_{12\text{deg}}[RII]$$

$$v_{13} = k_{13syn} - k_{13\text{deg}}[RI] \qquad v_{14} = k_{14syn} - k_{14\text{deg}}[S2_{cyt}]$$

$$v_{15} = k_{15syn} - k_{15\text{deg}}[S4_{cyt}] \qquad v_{16} = (k_{16\text{deg}} + k_{16lid})[R^C]$$

$$v_{17} = k_{17imp}[pS2_{cyt}] \qquad v_{18} = k_{18a}[pS2_{nuc}][S4_{nuc}] - k_{18d}[pS2S4_{nuc}]$$

$$v_{19}' = k_{19dp}[pS2_{nuc}] \qquad v_{20} = k_{20lid}[pS2_{nuc}]$$

$$v_{21} = k_{21\text{int}}[RII] - k_{21rec}[RII_{in}] \qquad v_{22} = k_{22\text{int}}[RI] - k_{22rec}[RI_{in}]$$

$$v_{23} = k_{23rec}[R_{in}^C]$$

$$\frac{d[RII]}{dt} = -v_1 + v_{12} - v_{21} + v_{23} \qquad \frac{d[TGF\beta : RII]}{dt} = v_1 - v_2 \qquad \frac{d[RI]}{dt} = -v_2 + v_{13} - v_{22} + v_{23}$$

$$\frac{d[R^C]}{dt} = v_2 - v_3 - v_{16} \qquad \frac{d[R_{in}^C]}{dt} = v_3 - v_4 + v_5 - v_{23} \qquad \frac{d[R_{in}^C : S2_{cyt}]}{dt} = v_4 - v_5$$

$$\frac{d[S2_{cyt}]}{dt} = -v_4 - v_{10} + v_{14} \qquad \frac{d[pS2_{cyt}]}{dt} = v_5 - v_6 - v_{17} \qquad \frac{d[pS2S4_{cyt}]}{dt} = v_6 - v_7$$

$$\frac{d[pS2S4_{nuc}]}{dt} = v_7 - v_8 + v_{18} \qquad \frac{d[S2S4_{nuc}]}{dt} = v_8 - v_9 \qquad \frac{d[S2_{nuc}]}{dt} = v_9 + v_{10} + v_{19}$$

$$\frac{d[S4_{nuc}]}{dt} = v_9 + v_{11} - v_{18} \qquad \frac{d[S4_{cyt}]}{dt} = -v_6 - v_{11} + v_{15} \qquad \frac{d[pS2_{nuc}]}{dt} = v_{17} - v_{18} - v_{19} - v_{20}$$

$$\frac{d[RII_{in}]}{dt} = v_{21} \qquad \frac{d[RI_{in}]}{dt} = v_{22}$$

The approach to obtain model parameter values is summarized as follows:

1. *Initial Rough Estimation*: Several kinetic parameter values were determined through an extensive literature search; some were computed using available *in vitro* experimental data; we also used physical constraints to determine others. The remaining unknown parameters were provided with initial estimates and reasonable upper and lower bounds by comparison with similar circumstances in the literature (e.g. similar steps in previous models or other signaling pathway models) and from known physical limitations (e.g. the diffusion-limited rates, 10^8-10^9 $M^{-1}s^{-1}$).

2. *Parametric Sensitivity Analysis*: To identify which parameters are the most important and which must therefore be estimated most precisely, using the set of initial estimates determined in Step 1 above, we performed local parameter sensitivity analysis to determine the effect of parametric changes on the set of five system responses of interest for which experimental measurements are available (i.e. total Smad2 in the nucleus and the cytoplasm[4], total phosphorylated Smad2 in the nucleus and the cytoplasm[4, 5], and total Smad4 in the nucleus[4]). The computations are based on the following expression for the normalized sensitivity coefficient (NSC):

$$NSC_{ij}(t) = \frac{p_j}{y_i}\frac{\partial y_i(t,\mathbf{p})}{\partial p_j}\bigg|_{\mathbf{p}}, \qquad i = 1,2, \ldots, 5; \qquad j = 1, \ldots, 37$$

where y and **p** respectively denote the system response variables and kinetic parameters. A total of 13 parameters were selected to be estimated more precisely because of their high sensitivity coefficients and/or because we had little or no confidence in their initial values.

3. *Least Squares Fitting to Data*: We fit our model predictions simultaneously to corresponding *in vitro* experimental data from the literature, via local minimization of the sum of squared residual errors:

$$\min_i R(\mathbf{p}) = \frac{1}{2}\sum (y_i(t,\mathbf{p}) - y_i^*(t))^2$$

where $y_i(t,\mathbf{p})$ and $y_i^*(t)$ denote, respectively, model predictions for a given trial of parameter values, **p**, and the corresponding experimental measurements, for each measured variable, *i.* The experimental data used for the curve-fitting are time courses of (i) total Smad2 in the nucleus[4], (ii) total Smad2 in the cytoplasm[4], (ii) total phosphorylated Smad2 in the nucleus[5], (iv) total phosphorylated Smad2 in the cytoplasm[4], and (v) total Smad4 in the nucleus[4].

4. *Identifiability*: We performed a "practical identifiability" analysis to determine whether the unknown parameters of the postulated model can be uniquely estimated from the available data, following Birtwistle *et al.*[6]. Briefly, approximate local confidence intervals for the parameter set are given by,

$$CI_i = t_{\alpha/2}^{N_t - N_p}\sqrt{\frac{S}{N_t - N_p}}\sqrt{(\mathbf{Z}^T\mathbf{Z})_{ii}^{-1}}$$

where N_t and N_p respectively denote the number of experimental data time points and the number of parameters to be estimated; $t_{\alpha/2}^{N_t - N_p}$ is the Student's *t*-distribution statistic evaluated with $N_t - N_p$ degrees of freedom, at confidence level 100(1- α)%, (with α as the "tail area probability" typically set at 0.05 to yield a 95% confidence level); *S* is the sum of squared errors, and ***Z*** is the model sensitivity matrix evaluated at the current parameter values. The i^{th} parameter is said to be *practically locally identifiable* only if the magnitude of its approximate confidence interval is less than a specified tolerance i.e. $|CI_i| < \varepsilon_i$.

5. *Identifiable Parameter Estimate Refinement*: Estimated values for identifiable parameters were further refined by repeating Step 4 (local identifiability test) followed by Step 3 (local least squares estimation). After obtaining the "best" estimates of this subset of parameters, we carried out a final least squares estimation of the entire parameter set.

Results and Discussion

To help understand quantitatively which aspects of the pathway most affect the system behavior, we carried out parameter sensitivity analysis for nuclear pSmad2-Smad4 complex, which is necessary for TGF-β-induced transcriptional regulation. Figure 2 shows normalized sensitivity coefficients as a function of time for the 10 most important parameters. The most important features of this plot are summarized as follows: (1) Immediately after ligand stimulation, the output variable is strongly affected by four of this set of most sensitive parameters: in order of importance, these are k_{4a} (binding of Smad2 to active receptors), k_{3int} (internalization of receptor complexes), k_{7imp} (nuclear import of pSmad2-Smad4), and k_{2a} (complex formation of activated TβRII and TβRI). (2) On the other hand, in the mid- to longer time interval after ligand stimulation, the following parameters become more important: in order of importance, these are k_{18a} (association of nuclear pSmad2 and Smad4), and k_{18d} (dissociation of nuclear pSmad2-Smad4), k_{11exp} (nuclear export of Smad4), k_{20lid} (ligand-induced degradation of pSmad2), k_{11imp} (nuclear import of Smad4), k_{23rec} (recycling of internalized receptor complexes), k_{3int} (internalization of receptor complexes), and k_{4a} (binding of Smad2 to active receptors).

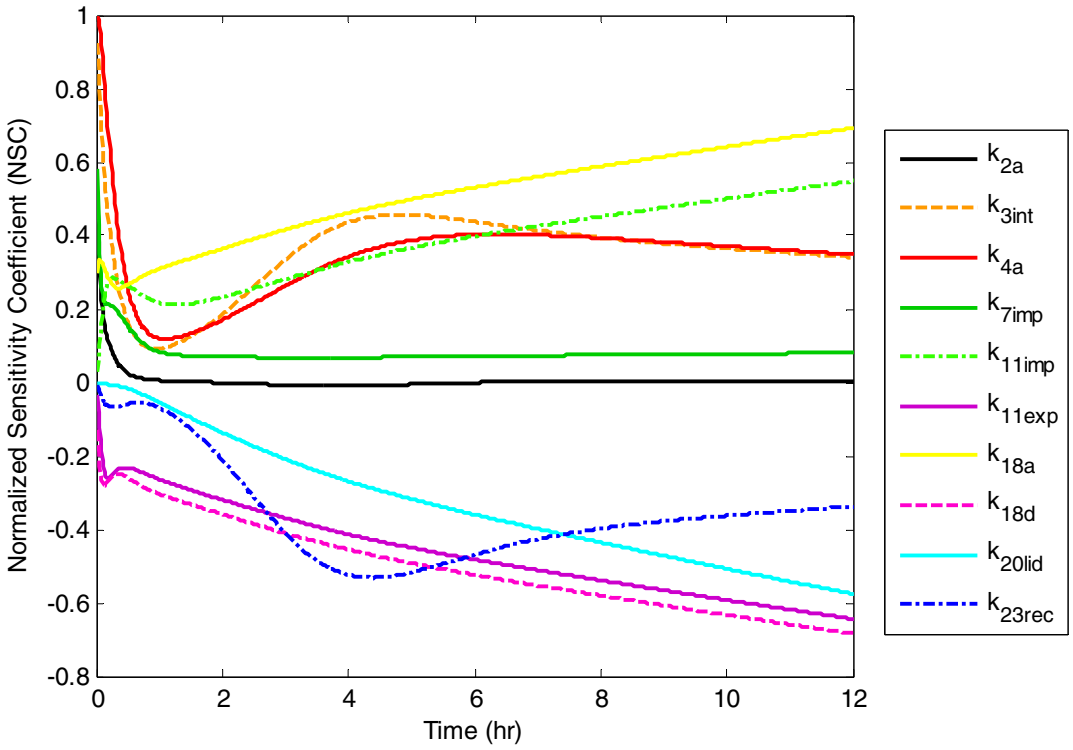

Figure 2. Model parameter sensitivities for select parameters with the greatest influence on phosphorylated Smad2-Smad4 complex in the nucleus.

These results have biologically important consequences. In particular, the increasing nature of the sensitivity coefficients of k_{18a} and k_{18d} over time shows that both the formation of pSmad2-Smad4 complex in the nucleus and its dissolution are crucial for nuclear retention of these complexes. To examine how variations in Smad complex formation in the nucleus

affects nuclear accumulation of activated Smad complexes, we varied the rate of association between nuclear pSmad2 and Smad4 (k_{18a}) 10-fold. Figure 3 shows that while rapid formation of the complex between nuclear pSmad2 and Smad4 induces prolonged and enhanced nuclear accumulation of pSmad2-Smad4 complex, slow binding of pSmad2 and Smad4 leads to shortened and attenuated retention of pSmad2 complex in the nucleus. Thus, these results imply that the nuclear complex formation step plays an important role in regulating the intensity and duration of TGF-β-targeted transcriptional activities through pSmad2 complexes. More importantly, the results imply that pSmad2 complex-mediated response to TGF-β stimulation may be significantly attenuated by competitive inhibition or by interference from other nuclear molecules that also have high affinity for either pSmad2 or Smad4. This inhibitory action ultimately gives rise to a significant reduction in the rate of association between these proteins. Taken together, these results reveal that the step of complex formation between pSmad2 and Smad4 is closely associated with modulation of TGF-β-induced signal patterns.

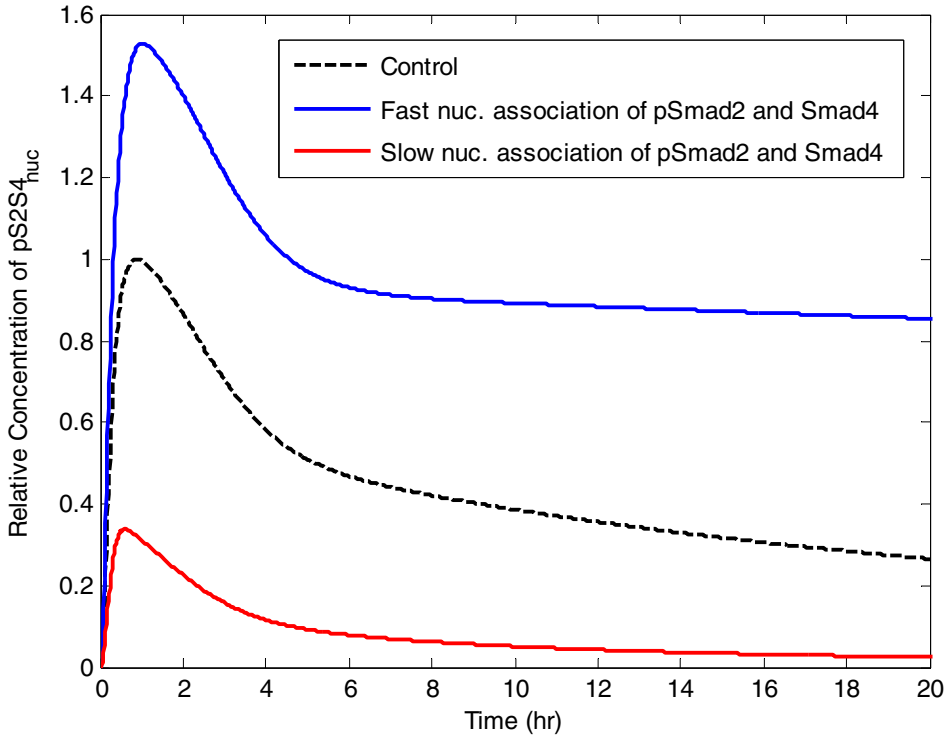

Figure 3. The effect of variations in the rate of nuclear pSmad2-Smad4 association

To help understand how cancer cells can become resistant to the tumor-suppressor effects of TGF-β, but, at the same time, remain responsive to the tumor-promoter effects, we investigated the effect on the TGF-β signaling system of abnormal alterations (e.g. mutations, deletions, downregulation, etc.) in receptors, using a 10-fold reduction in the initial levels and production rate of both Type I and Type II receptors to represent cancerous conditions. First, Figure 4 indicates that the amount of TGF-β needed to produce a saturated Smad-mediated response in cancer cells is far higher than that in healthy cells. Specifically, while the response for normal cells is essentially saturated with 1 pM of TGF-β (with higher doses producing essentially the same response), at least 10 pM of TGF-β is required before the Smad-mediated response begins to approach saturation. This is, of course, a direct effect of the reduction in

the number of functional receptors in cancer cells which renders them less responsive to TGF-β stimulation. But this finding also indicates an important characteristic of cancerous cells: to elicit nuclear Smad-mediated activity generally requires *more* TGF-β than normal.

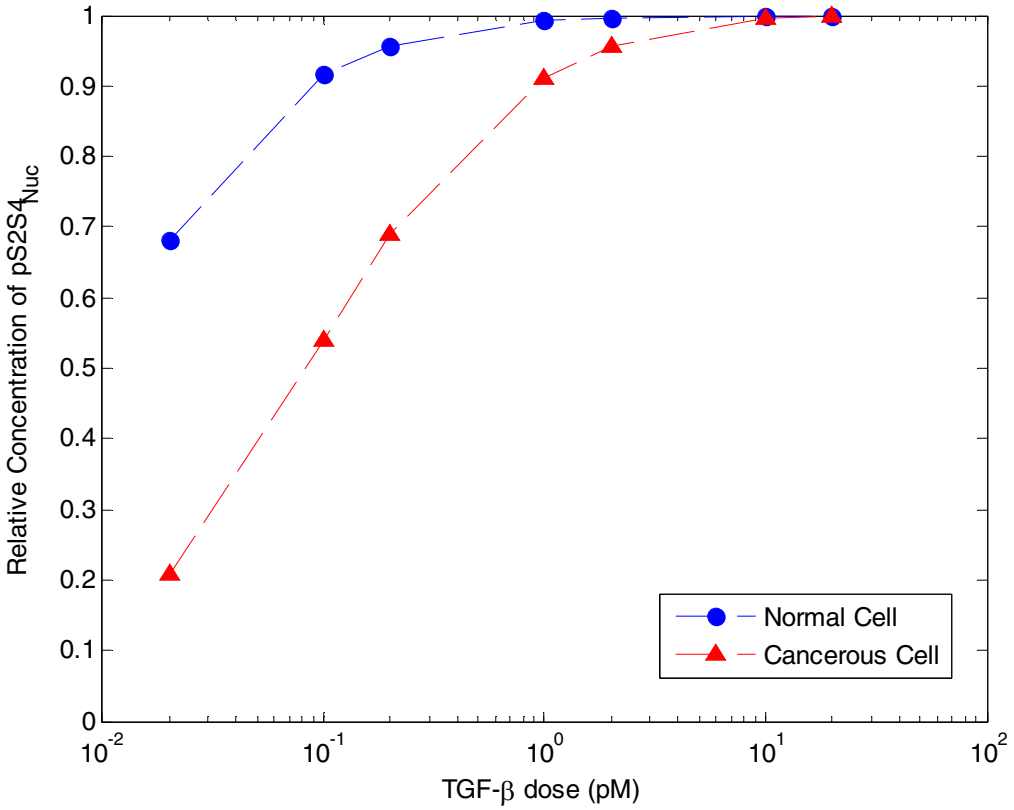

Figure 4. The effect of different concentrations of TGF-β on the maximum responses of activated nuclear Smad2-Smad4 complex in normal cells (blue circles) and in cancerous cells (red triangles).

Next, a head-to-head comparison of normal versus cancerous cell responses reveals that the sharp drop in the level of functional receptors in cancer cells leads to a marked decrease in the activity of nuclear pSmad complexes (Figure 5). Compared to the normal cell response, the peak activity of nuclear pSmad complexes in cancer cells was reduced by 65%, with the steady-state activity also remaining comparatively low. Interestingly, a reduction in the level of receptors also slowed nuclear pSmad responses. While nuclear pSmads in the normal system reached maximum activity in 55 min, their activity under a cancerous condition peaked at 86 min. Taken together, these results indicate generally that a reduction in the level of functional TGF-β receptors in cancer cells may lead to attenuated and slower TGF-β-stimulated signaling responses via Smad2. The specific implications of the model predictions in Figures 4 and 5 reveal some potentially important findings about TGF-β and cancer cells: (i) cancer cells require *higher* than normal levels of TGF-β in order to elicit significantly attenuated (and much slower) nuclear Smad-mediated activity; (ii) but even the increased levels of TGF-β will never be able to produce Smad-mediated responses that will be anywhere close to normal because of the saturation effect.

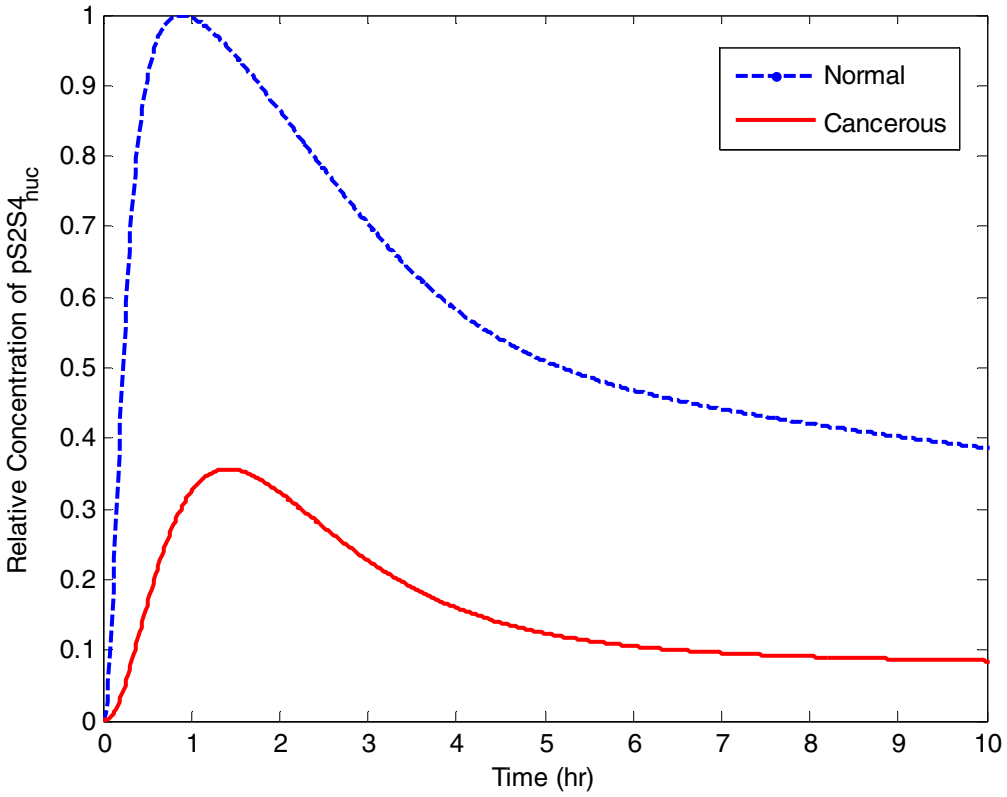

Figure 5. *In silico* mutation results: Responses of nuclear pSmad-Smad4 complex to 10-fold reduction in initial levels and protein synthesis rate constants of both Type I and Type II receptors.

As shown above via simulation, cancer cells may have attenuated TGF-β-stimulated Smad pathway responses. Cancer cells have been confirmed experimentally to be resistant to the antiproliferative effect of TGF-β, while showing typical pro-oncogenic responses. Such behavior may be explained in part by the following "threshold hypothesis": in response to TGF-β, growth-inhibitory genes require higher threshold levels of nuclear Smad activity for their expression than genes associated with pro-oncogenic and pro-metastatic effects. In other words, under normal conditions, or in the early stage of cancer progression, the antiproliferative responses to TGF-β are predominant over pro-oncogenic responses. This is because the transcriptional activity of nuclear pSmad is high enough to induce anti-growth gene expression. However, as cancer progresses, this transcriptional activity may decline significantly and thereby hardly exceed the threshold necessary for the expression of growth-inhibition genes. Meanwhile, genes related to tumor-promoting effects may be relatively insensitive to the attenuation of the transcriptional activity by Smads, so that the expression of such genes remains approximately unchanged even under cancerous conditions. As a consequence, the dominance of tumor suppressor genes over the tumor-promoter genes may be blunted in cancer cells. We believe that further investigation into differences in the temporal profiles of gene expression and thresholds of anti-growth and pro-oncogenic genes induced by TGF-β will provide some clues regarding the putative dual effects of TGF-β.

Conclusion

In this work we have presented a comprehensive computational model of the TGF-β signaling pathway that describes how an extracellular signal of TGF-β ligand is sensed by receptors and transmitted into the nucleus through intracellular Smad proteins. The model yields quantitative insight into how TGF-β-induced responses can be modulated and regulated. The model also allows us to predict possible dynamic behavior of the Smad-mediated pathway in abnormal cells, and provides clues regarding possible mechanisms for explaining the seemingly contradictory roles of TGF-β during cancer progression. We believe that incorporation of the effect of crosstalk among other important signaling cascades and detailed gene expression mechanisms into the current model will facilitate better understanding of the TGF-β paradox in cancer.

References

1. Shi, Y., and J. Massague. 2003. Mechanisms of TGF-β signaling from cell membrane to the nucleus. Cell 113:685-700.
2. Feng, X. H., and R. Derynck. 2005. Specificity and versatility in TGF-β signaling through Smads. Annu Rev Cell Dev Biol 21:659-693.
3. Pardali, K., and A. Moustakas. 2007. Actions of TGF-β as tumor suppressor and pro-metastatic factor in human cancer. Biochim Biophys Acta 1775:21-62.
4. Pierreux, C. E., F. J. Nicolas, and C. S. Hill. 2000. Transforming growth factor β-independent shuttling of Smad4 between the cytoplasm and nucleus. Mol Cell Biol 20:9041-9054.
5. Inman, G. J., F. J. Nicolas, and C. S. Hill. 2002. Nucleocytoplasmic shuttling of Smads 2, 3, and 4 permits sensing of TGFβ receptor activity. Mol Cell 10:283-294.
6. Birtwistle, M. R., M. Hatakeyama, N. Yumoto, B. A. Ogunnaike, J. B. Hoek, and B. N. Kholodenko. 2007. Ligand-dependent responses of the ErbB signaling network: experimental and modeling analyses. Mol Syst Biol 3:144.