行政院國家科學委員會專題研究計畫 期中進度報告

利用系統生物學策略建立定量整合性的雌激素作用模式——
總計畫(2/3)
期中進度報告(精簡版)

計畫類別：整合型
計畫編號：NSC 98-2627-M-006-005-
執行期間：98年08月01日至99年07月31日
執行單位：國立成功大學化學系（所）

計畫主持人：陳淑慧

處理方式：本計畫可公開查詢

中華民國 99年06月21日
**Introduction**

The main project is aiming to examine estrogen action by integrated proteomics and informatics approach. 17β-Estradiol (E2) is an ovarian hormone necessary for the development of secondary sexual characteristics and function of the reproductive system in females. It also plays important roles in non-reproductive organs by its antioxidant activities. E2 action involves genomic pathway through classical cytosolic/nuclear receptors acting as nuclear transcriptional factors and nongenomic pathway mediating rapid effects through unusual membrane receptor or classic cytosolic/nuclear receptor spanning through plasma membrane or acting through multi-protein complex associated to the inner part of plasma membrane. In genomic pathway, estrogen receptor α (ERα) will become dimer under E2-treatment and translocate into nucleus to bind with estrogen response element (ERE) and activate downstream transcription actions.

In subproject I, we developed an integrated proteomics approach using chemically functionalized gold nanoparticle (AuNP) as a novel probe for affinity purification of large protein complex *in-vivo* and applied this approach to globally map the activation complexes of estrogen response element (ERE) and 17β-estradiol (E2). We term such approach as Quantitative Nano-Proteomics for resolving protein complexes (QNanoPX). Two kinds of probes, AuNP-ERE and AuNP-E2 were fabricated by self-assembly monolayer technique and the affinity pull down was optimized for ERα protein from MCF-7 cells. We discovered some novel complex proteins such as cMyc and TIF1β that may play a role in the transcriptional action of ER. Two major tasks were accomplished in Subproject II. Task 1 was to explore the role of proteins found in Subproject I using the technology of AuNP-ERE pull-down assay such as TIF1β. Task 2 was to explore the role of ERβ in tumor development. Since Task 1 was to address the role of ERα in cell proliferation, MCF-7 were used in Task 1. In contrast, Task 2 was to explore the role of ERβ in cell proliferation, Hela were used in Task 2. In both tasks, transfection with shRNA and over-expression of ERβ were used to modulate the expression of ERβ and the proteins identified. Estrogen response element (ERE) reporter assay was developed to confirm if the protein identified elicited cell proliferation was ERE dependent. In addition, immunoproteomic analysis was used to discover ERβ-associated proteins. In subproject III, we developed a protein surface distillation and comparison method to predict the DNA binding proteins which have similar or dis-similar binding domains as compared to the positive and negative controls. We employed the Voronoi diagram and Delaunay triangulation techniques to distill the amino acids which locate on the
protein surface. The distilled amino acids forms several triangle plans to represent the surface entities. Also, two matrices - Amino Acid Substitution Matrix and Metric SSE Exchange Matrix were combined with RASA (relative solvent accessible area) to be further applied to identify surface features. Finally, a surface comparison method is applied to compare whether there are high structure similarity triangle planes, and to assess if the two proteins exist high resembling surfaces.

Two proteins are similar in steric structure if the amino acid sequence of the fragment of one protein is near 80% identical to the amino acid sequence of the fragment of the other protein. In order to determine the structure of the fragment of an unknown protein, it is reasonable to choose the fragment having similar sequence of an experimentally structure-known protein as a model. One then predict and establish the structure of the fragment of the unknown protein, which is under study to meet research interests. This approach is termed Homology modeling methodology. Conversely we can search for proteins of similar functions based on their structural analysis. It is therefore promising to search for proteins capable of interaction with ERE through the major complexation part of the adduct of ER-alpha and ERE. It is noticeable that the results are intrinsically limited since they are totally dependent upon results having been reported. Furthermore, the driving force of interactions involved in formation of these supramolecules still remains unknown. One cannot extract the information concerning specific mechanisms for specific purpose of these complexes of proteins. Viewing these needs to expand and advance our research results, Prof. Wang (王晓萍), coPI of the main project, are exploring the chemically driving force or the chemical origin leading to the complex of ER-alpha and ERE from its structural network. Knowing the microscopic nature of this complex by molecular orbital calculations, it is expected to understand its physical properties. These results will not only provide the origins for complexation of the two proteins but serve as adjusting parameters for modifying models to be established to get more precise results in the sub-research stated earlier.

Methods

To establish communication channels among main and subprojects, regular meetings were held biweekly to share results and information as well as to discuss future works. A brief list of meeting times and topics are in below.
Classification of the AuNP-ERE pulled down large-scale proteomics result.

1. Use the structure prediction software developed in first year to calculate the large-scale proteomics result.
2. ERE luciferase plasmid was designed and ordered.

8 positive and 8 negative proteins were used as training model and the structure prediction software was used to test these known proteins. The result showed that 15 proteins were predicted correctly, but only one positive protein was predicted as negative.

1. Used references to check the structure prediction result.
2. Find out the correlation of cMyc, TIF1b and ERα after TIF1b siRNA treatment.

1. Comparison of three computational methods and tried to add in hydrogen atoms.
2. Find out the association between molecular function, biological process and cellular component in text mining results.

E2 treatment will increase S phase and inhibit G0/G1 phase.

The structure prediction software was used to score the AuNP-ERE pulled down proteins and the result showed that seven proteins were predicted positive.

Recombinant ERα and cMyc were pulled down by AuNP-ERE. The amount of pulled down ERα decreased under the present of cMyc.

The correlation between ERα, cMyc and ERβ under 15 min, 3 h and 24 h E2 treatment was investigated.

Time course of cMyc expression in MCF-7 cells under E2 treatment was studied.

Data mining will use pathway interaction database to find out the transcription relations following classification of transcription factor and target gene using gene ontology terms.

Protein (ERα, cMyc and ERβ) expression of E2 treatment under different time course showed same trends of ERα and cMyc.

Text mining result revealed that there might be one or two layers between ERα
and cMyc.

20100119

1. E2 involves the protein synthesis of cMyc but not TIF1b.
2. If 8 positive and 8 negative were used as training dataset, accuracy will be 93.8%.
   If 8 positive and 232 negative were used, accuracy will be 99.6%.

20100205

Both ERα and cMyc have the highest expression under 15 min E2 treatment.

20100302

1. The binding of TIF1b on the ERα-ERE complex may suppress the expression of cMyc and enhance the expression of ERα.
2. It is still not clear if TIF1b is associated with ERα.
3. TIF1b depletion increases cell death which is E2 dependent. The data suggest TIF1b may be an important adaptor, which stabilize protein complex to maintain cell survival.

20100413

1. E2 will lose its transcription activity after TIF1b depletion under E2 treatment.
2. Our text mining algorithm could predict unknown, but potentially exist ERα and cMyc interaction relation from current publications.

20100511

1. From recombinant ERα and cMyc experiment, the amount of cMyc pulled down by AuNP-ERE probe increased with increasing existence of ERα.
2. The luciferase activity showed highest intensity under 24 h E2 treatment.

20100525

Five different datasets (BioGRID, DIP, HPRD, MINT and Reactome) were used to modify the text mining algorithm.

20100608

14-3-3 beta/alpha protein were pulled down by AuNP-ERE probe from in-vivo experiment and also predicted to have high score in our structure prediction software. To verify this result, recombinant 14-3-3 beta/alpha protein was purchased and pulled down by AuNP-ERE probe. The positive result confirms the accuracy of our model.
Results and discussion

Required key platforms have also been well conducted by each subproject as outlined in below.

Subproject I

The AuNP-ERE and AuNP-E2 probes have been fabricated and their capability for affinity capture of ERα from MCF-7 cells has been investigated and both show excellent binding specificity and capacity. QNanoPX approach using AuNP-ERE was attempted and a total of 304 proteins were identified and quantified from the pull down. A majority of identified proteins with significant binding are involved in genomics pathway including known and unknown coactivators/corepressors and a manuscript is currently under submission. We have also completed the design and synthesis of AuNP-E2 probe as well as its pull-down assays for microscale (μg) protein purification. For this synthesis, the precursor molecule, 17α-substituted estradiol derivative with an amine linker (E2-NH₂), was synthesized and bound to a biotin-NHS molecule to form biotin conjugated E2 (E2-biotin) probe. On the other hand, Au nanoparticle (AuNP) synthesized by citrate reduction was modified with avidin and polyethylene glycols (PEG). The avidin functionalized AuNP (AuNP-avi) was characterized to have a hydrodynamic diameter around 36 nm and each AuNP was characterized to have around 50 bound avidin molecules. The biotin-E2 probe was used to capture estrogen receptor α (ERα) from MCF-7 cells and the captured proteins were subsequently pulled down by AuNP-avi. The recovery yield of ERα was determined to be higher than 90% and the enrichment factor of the probe was 4.8-fold higher than the control probe.

Subproject II

In subproject II, the technology of transfection of shRNA was used to explore the role of TIF1beta in the proliferation of MCF-7 and the relation of cMYC with E2-mediated cell proliferation. To further explore whether TIF-1beta elicited cell proliferation via its binding to estrogen response element (ERE), ERE reporter assay was used. Our data clearly showed that the expression of TIF1beta was positively correlated with the expression of ERα. The deletion of TIF1beta enhanced the response of MCF-7 to E2-mediated cell cycle progression and E2-induced reporter activity. Since immunoprecipitation analysis did not show the increased association of TIF1beta with ERα, our data suggest that AuNP-ERE pull-down assay indeed discovers novel proteins which are associated with ERα. The protein associated with ERα indeed plays a functional role in E2-induced cell proliferation. However, the increase may be due to the mechanism of E2-induced post-translational modification.
other than the increase of protein abundance in E2-ER-mediated complex.

Subproject III

Among this research, a new method of protein alignment had been brought up. Experimental data showed that there existed upper than 2 times discrepancy of surface information between ERE (estrogen response element) and non-ERE binding proteins. Furthermore, as we concentrated on aligning eight major types of protein prone to bind with DNA, we discovered there was also upper than 7 times difference between the two groups. Finally, we proceeded to apply these results to ERE transcription factor predictions. Upon this part of research, classifiers increased 6.5% of average precision and 7.1% of average recall.

We used the estrogen response element (ERE) which has a consensus sequence of GGTCAGAGTGACC as a model and 7 binding transcription factors deduced from TRANSFAC including estrogen receptor α (ERα) as the positive control and another 7 proteins which were discriminated by TRANSFAC as the negative control to train the system we developed. Experimental data showed that there existed more than 2 times discrepancy of surface information between ERE positive and non-ERE negative binding proteins. Furthermore, as we concentrated on comparing eight major types of proteins prone to bind with ERE, we discovered there was also more than 2 times difference between these two groups. We then used the trained algorithm to quest for ERE binding proteins among 195 proteins in MCF-7 which structure could be deduced from protein databank (PDB). The classifier built with surface information increased 5% of average precision and 5.6% of average recall. Furthermore, we proved that six out of seven predicted ERE binding proteins could be identified from proteomics experiments via a pull-down assay using an affinity ERE probe (Molecular & Cellular Proteomics, 2010, 9, 209-224).

Limited by the size of computer memory, it is time consuming or even in vain to rely on ab initio calculation only. In the ER-alpha/ERE complex, for example, there are 1449 atoms contained in the A chain of ER-alpha and the bonded fragment of ERE (obtained from 1HCQ data of RCSB PDB). In the DFT calculations using STO-3G basis set for geometry optimization, there will be 4607 basis functions involved in the calculation. For the 4-cluster desk-top computer, this job is far beyond its capability. We are employing the QM/MM approach to resolve this problem. The macro-structure is divided into two parts. The first part dealing the primary structure involved in interaction is calculated by QM (Quantum Mechanics) and the rest is
The design of the whole calculational procedures is described as follows.

1. The targeted structure along with properties to be discussed is first extracted from protein data bank and followed by making minor structural modification to meet our needs. For example, The pdb parameters obtained from x-ray spectroscopy is needed for H-atom inclusion, which is pH-value dependent.

2. Further modification is performed by the Amber MM calculation package for geometry optimization, followed by monitoring the thermodynamic structure to get the initial structure for subsequent ab initio calculations.

3. The ONIOM (Our own N-layered Integrated molecular Orbital and molecular Mechanics) method supplied by Gaussian 03 package is employed for final structural optimization. The initial structure is divided into levels. The higher level is related the structure concerning interaction to be investigated and will be treated by QM calculations. The rest of the initial structure obtained at step (2) is classified as lower level and will be treated by MM calculations. Having assigned levels and input relevant calculation parameters, the final structure with electronic properties obtained by C-matrix elements are analyzed to extract informative and/or instructive information. The initial structure can be assigned as many as three levels depending on computing time and of the results in ONIOM calculations.

At the moment, we are making choice of proper computation package by comparison of results obtained from simplified systems. Detailed procedures such as creating necessary files for input and output data instructed by the manual are in thorough test.