Vol. 139, No. 1, 1986 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
August 29, 1986 Pages 281-286

COMPUTER-BASED SEARCH FOR STEROID AND DNA BINDING SITES ON ESTROGEN AND GLUCOCORTICOID RECEPTORS

Michael E. Baker*

Department of Medicine, M-023
University of California, San Diego
La Jolla, California 92093

Received June 11, 1986

SUMMARY. The primary amino acid sequences of proteins that are receptors for estrogen, glucocorticoids, and ouabain were compared with each other using computer programs designed to detect and quantify similarities between proteins. Three regions of similarity between the estrogen receptor (ER) and the glucocorticoid receptor (GR) were identified. On the ER, residues 173-250, 323-395, and 426-458 are similar to residues 409-486, 540-612, and 644-676, respectively, on the GR. The ALIGN computer analysis of these segments on the ER and the GR gave comparison scores that were 16.8, 13.7, and 6.8 standard deviations higher, respectively, than that obtained with a comparison of randomized sequences of these proteins. The probability of getting these scores by chance is less than 10^{-60}, 10^{-40}, and 10^{-11}, respectively. Others have proposed that the segment on the ER and GR that is nearest their amino terminus (e.g. residues 173-250 of the ER) is part of their DNA binding domain and that the other two similar segments on each receptor, which are closer to their carboxy terminus, are part of their steroid binding domain. Here, we present evidence to support both of these hypotheses. First, an Align computer analysis indicates that residues 323-395 of the ER and residues 570-612 of the GR contain a region that is similar to a part of the α-subunit of the $(Na^+ + K^+)$ATPase that is hypothesized to bind the steroid ouabain. This similarity provides additional support for the proposed location of the steroid binding site on the ER, GR, and $(Na^+ + K^+)$ATPase. Second, a computer search of the protein sequence database revealed that protamine, a DNA binding protein, has some similarity to residues 255-281 of the ER, which are thought to be part of the DNA binding domain in the ER. Further, we find that residues 276-281 of the ER contain a structure that has been found at the nucleotide binding domain of some protein kinases. If this region on the ER binds ATP, then it may be involved in phosphorylation/dephosphorylation of the ER, which is thought to be important in its mechanism of action.

Certain general principles concerning the mechanism by which steroid hormones stimulate the transcription of genes in target tissues are understood (1-3). Steroids first bind to a specific intracellular protein (receptor), which then undergoes a conformational change that increases its affinity for DNA and/or its associated proteins. The binding of this form of the receptor to a nuclear site(s) stimulates the synthesis of various proteins. However, the details of this process, such as: what hydrophobic and polar properties of the steroid and its binding site on the receptor influence their association, how this association changes the conformation of the receptor so as to increase its affinity for nuclear sites; and how the receptor binds to the nuclear site sites are only beginning to be elucidated. The recent reports of the primary amino acid sequences of the glucocorticoid receptor (GR) (4,5) and the estrogen receptor (ER) (6,7) provide an essential first step in understanding those details of steroid mediated

* Part of this work was done under the auspices of the Visiting Scientist Program at the National Biomedical Research Foundation, Georgetown University Medical Center, Washington, D.C.

0006-291X/86 $1.50

processes. Additional information about the steroid binding site on proteins could come from an analysis of the primary amino acid sequence of the the α-subunit of the $(Na^+ + K^+)$ATPase (8,9), which binds the steroid ouabain. This protein is an integral membrane protein that couples the hydrolysis of ATP to the transport of Na^+ and K^+ across the plasma membrane. The ouabain binding site is on the extracellular part of the protein, which is helpful for localizing the steroid binding site because very little of the protein is extracellular. Based on several criteria, Shull et al. (8) hypothesized that part of the ouabain binding site was on residues 307-312 of this protein. To seek clues to the location of the steroid binding site on the GR and ER, as well to evaluate Shull et al's hypothesis, a computer-based comparison of the amino acid sequences of the GR, ER, and the α-subunit of the $(Na^+ + K^+)$ATPase was performed. As reported here, these analyses show that a segment on the ER and the GR that is thought to contain their steroid binding site (4-7) is similar to the proposed ouabain binding site of the ATPase (8). We also find that part of the proposed DNA binding region on the ER (6,7) has some some resemblence to protamine, a DNA binding protein. Our findings may be useful in defining sites for mutagenesis studies for understanding the functions of different domains of steroid receptors.

METHODS. A search for similarities between either the ER or the GR and the proteins in the database and was done using the program of Lipman and Pearson (10). Initial comparisons of protein sequences were done using either the graphical DOTMATRIX program (11,12) or the RELATE program (13,14). Protein segments that appeared to be similar were then analysed further using either the RELATE program or the ALIGN program (13).

The RELATE program compares all possible segments of a given length from one sequence with all segments of the same length from a second sequence. A segment score is accumulated from the pair scores of the amino acids occupying corresponding positions with the two segments. The pair scores are specified using an empirically derived mutation data matrix. The mean of a number of highest scores is determined for the given sequences and for 1,000 comparisons of random permutations of the sequences. The segment comparison score is calculated as the difference between the mean of the real sequences and the averaged value determined from the randomized sequences divided by the standard deviation (SD) of the values of the randomized sequences. The segment comparison score is thus expressed in SD units. A score greater than 5 SD units (p less than 2.8×10^{-7}) is usually interpreted as indicating that the sequences of the two proteins are similar, which suggests that the proteins may also have similarities in their biological properties. Scores between 3 SD ($p < 10^{-3}$) and 5 SD ($p < 2.8 \times 10^{-7}$) are less conclusive about a similarity in properties between the two proteins. In this case, the information from the computer analyses needs to be combined with other evidence before reaching a conclusion about the similarities and differences in the properties of the two proteins.

The ALIGN program calculates the best alignment between any pair of sequences given a matrix of amino acid pair scores and a penalty for breaking a sequence (gap). This score is compared with that obtained from 2,500 random permutations of the two sequences. The alignment score is the number of standard deviations by which the maximum score for the real sequences exceed the average maximum score for the random. For the analyses reported here, the mutation data matrix was used with a bias of 6 and a gap penalty of 6.

RESULTS

A. *Comparison of estrogen and glucocorticoid receptors.* A DOTMATRX analysis of the human estrogen and glucocorticoid receptors indicated that three regions in these proteins appear to be similar.

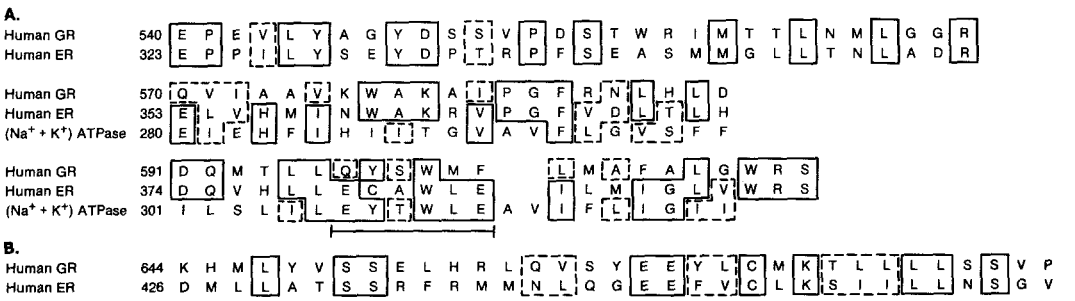

Figure 1. Alignment of the human estrogen and human glucocorticoid receptors and the α-subunit of the sheep $(Na^+ + K^+)$ATPase.

A) Alignment of residues 323-395 of the human ER with residues 540-612 of the human GR. Also shown is an alignment of residues 280-321 of the α-subunit of the sheep $(Na^+ + K^+)$ATPase with resides 353-395 of the human ER and with residues 570-612 of the human GR. In this segment, *Torpedo californica* $(Na^+ + K^+)$ATPase differs from that of the sheep by only one residue. *Torpedo californica* has a glycine residue instead of the glutamic acid residue at position 307 of the sheep $(Na^+ + K^+)$ATPase. Solid boxes show identities and the dotted boxes show conservative replacements according to the scheme: (P,G), (M,C), (Y,W,F,H), (L,V,I,A), (K,R), (E,Q,N,D), and (S,T). The ALIGN comparison of the GR and ER yields a score of 13.7 standard deviation units. The ALIGN comparison for the human ER and sheep ATPase yields a score of 6.2 standard deviation units. A gap of 2 amino acids has been inserted in both the ER and the GR to improve their alignment with the $(Na^+ + K^+)$ATPase. The proposed ouabain binding segment is underlined.

B) Alignment of residues 426-458 of the human ER with residues 644-676 of the human GR. The ALIGN comparison yields a score of 6.8 standard deviation units.

Residues 173-250, 316-395, and 426-458 of the ER are similar to residues 409-486, 532-612, and 644-676, respectively, of the GR. The similarity was quantified using the ALIGN analysis. The comparison scores were 16.8, 13.8, and 6.8 standard deviations higher than were obtained with a comparison of randomized sequences of these segments. The probability of obtaining these scores by chance is less than 10^{-60},10^{-40}, and 10^{-11}, respectively. This similarity suggests that these segments on the ER and the GR have common functions. The alignment of the segments near the amino terminus has been report by Greene et al(7). Figure 1 shows the alignment of the other two segments.

B. *Comparison of the α-subunit of the $(Na^+ + K^+)$ATPase with estrogen and glucocorticoid receptors.* The $(Na^+ + K^+)$ATPase is an integral membrane protein that binds the steroid ouabain. The ouabain binding site is on the extracellular part of this protein. A comparison of the α-subunit of the $(Na^+ + K^+)$ATPase with the GR and the ER indicated that part of these proteins were similar. The ALIGN analysis of residues 280-321 of the $(Na^+ + K^+)$ATPase with residues 353-392 of the ER and residues 570-612 of the GR gave scores that were 6.2 and 5.9 standard deviations higher, respectively, than that obtained with 2,500 randomized sequences of these proteins. The probability of getting these scores by chance is less than 10^{-9}. Figure 1a shows the alignment of these segments.

Figure 2. Alignment of residues 255-281 of the human estrogen receptor with sturgeon protamine. Solid boxes show identities and the dotted boxes show conservative replacements. Residues 276-281 have the structure Gly-X-Gly-X-X-Gly that has been found at the nucleotide binding site of some protein kinases and dehydrogenases.

C. *A similarity between the estrogen receptor and protamine?* A search of the protein sequence database revealed an interesting similarity between residues 255-281 of ER and the sequences of certain sturgeon (15,16), trout (17), and herring (18) protamines, which are proteins that bind to DNA. Figure 2 shows the alignment of this region of the ER with a sturgeon protamine (15). There are numerous identities and conservative replacements in these two segments on the ER and protamine. The similarity between protamine and the ER was quantified using the RELATE analysis. The comparison score was 4.4 standard deviations higher than that of 1,000 randomized comparisons of these segments. The probability of getting this score by chance is about 3×10^{-6}.

DISCUSSION A. *Location of the steroid binding site.* The similarity between GR and ER in the region shown in Figure 1 suggests a similar function for this region in these proteins. Supporting this suggestion is the similarity in the size of the gaps between these segments on the ER and the GR, which would permit the two segments on each receptor to be spatially aligned in the native receptor species. Structural studies have indicated that the steroid binding site on the GR resides on its carboxy terminal half (4,5); the segments shown in Figure 1 may contain the steroid binding site. Evidence to support this hypothesis and even to localize where part of the interaction of the steroid with the receptor occurs comes from the similarity between the ER, GR, and $(Na^+ + K^+)$ATPase revealed in Figure 1a because this region contains residues 307-312 of the ATPase, which have been proposed to be an essential part of the ouabain binding site (8). The similarity in this small segment includes a tryptophan that is thought to be important in the binding of ouabain to the $(Na^+ + K^+)$ATPase (8). The ALIGN comparsion of the ER and $(Na^+ + K^+)$ ATPase, which yields a score of 6.2 standard deviation units for the region shown in Figure 1a, provides quantitative support for the hypothesis that this is a conserved segment in these three steroid binding proteins. If this segment is indeed part of the steroid binding site on the ER and GR, then these similarities support Shull et al.'s hypothesis about the location of the ouabain binding site on the $(Na^+ + K^+)$ ATPase (8).

The erythroblastosis virus oncogene protein $p75^{gag-erb-A}$ (19) is homologous to both the GR (5) and the ER (6,7). While the homology is strongest in the amino terminus region of the GR and the

ER, residues 353-392 of the ER and residues 570-609 of the GR are also similar to part of v-erb-A (5,7). We find that an ALIGN comparison of these residues of the ER and the GR with residues 209-248 of v-erb-A yields scores that are 6.5 and 7.1 standard deviations higher, respectively, than that of randomized sequences of these segments. If these residues on the GR and the ER are part of their steroid binding domain, then this raises the possibility that v-erb-A could bind steroid-like molecules. With this in mind, we compared v-erb-A with the α-subunit of the sheep and *Torpedo californica* $(Na^+ + K^+)$ATPase using the RELATE program. No convincing similarities between these proteins was revealed. Thus, the computer analyses are ambiguous about the possibility that v-erb-A binds steroid-like ligands.

B. *Location of the DNA binding site.* The alignment of protamine, a DNA binding protein, with a region of ER that is thought to contain the DNA binding site is intriguing. This part of the ER comes just after a region that is highly conserved among the ER, the GR, and v-erb-A (5-7). Interestingly, RELATE and ALIGN computer analyses show that residues 491-517 on the GR, which would be expected to correspond to residues 255-281 of the ER, are not similar to either protamine or, for that matter, to residues 255-281 of the ER. If residues 255-281 of the ER are indeed part of its DNA binding domain, then this difference between the ER and GR could be important in the recognition of the specific nuclear binding site.

Finally, we note that residues 276-281 of the ER, which are the last 6 residues in the segment shown in Figure 2, have the structure Gly-X-Gly-X-X-Gly, where X can be an arbitrary amino acid. This structure has been found at the nucleotide binding domain of some protein kinases and dehydrogenases (14,20-22). Also, protein kinases contain a lysine residue 15 to 22 residues downstream from this glycine-rich sequence. This lysine interacts with ATP. Lysine 299 on the ER is in a similar position relative to glycine 281. If this region on the ER binds ATP, then it may be involved phosphorylation/dephosphorylation of the ER, which is thought to be important in its mechanism of action (23,24).

ACKNOWLEDGEMENTS. I thank the staff of the National Biomedical Research Foundation for advice and discussion during my stay there. I thank Darrell Fanestil for his interest in this research and the lively discussion that has resulted from that interest. I am grateful to Marvin F. Weissberg his continuing encouragement and support.

REFERENCES

1. Ringold, G.M. (1985) Annu. Rev. Pharmac. Tox. 25,529-566.
2. Katzenellenbogen, B.S. (1980) Annu. Rev. Physiol. 42, 17-35.
3. Gorski, J. and Gannon F. (1976) Annu. Rev. Physiol. 38, 425-450.

4. Hollenberg, S.M., Weinberger, C., Ong, E.S., Cerelli, G., Oro, A., Lebo, R., Thompson, E.B., Rosenfeld, M.G., and Evans, R.M. (1985) Nature 318, 635-641.
5. Weinberger, C., Hollenberg, S.M., Rosenfeld, M.G., and Evans, R.M. (1985) Nature 318, 670-672.
6. Greene, G.L., Gilna, P., Waterfield, M., Baker, A., Hort, Y., and Shine, J. (1986) Science 231, 1150-1154.
7. Green, S, Walter, P, Kumar, V., Krust, A., Bornert, J-M, Argos, P., and Chambon, P. (1986) Nature 320, 134-139.
8. Shull, G.E., Schwartz, A., and Lingrel, J.B. (1985) Nature 316, 691-695.
9. Kawakami, Kiyoshi, Noguchi, S., Noda, M., Takahashi, H., Ohta, T., Kawamura, M., Nojima, H., Nagano, Kei, Hirose, T., Inayama, S., Hayashida, H., Miyata, T., and Numa, S. (1985), Nature 316, 733-736.
10. Lipman, D.J. and Pearson, W.R., (1985) Science 227, 1435-1441.
11. Maizel, J.V. and Lenk, R.P. (1981), Proc. Natl. Acad. Sci. USA 78, 7665-7669.
12. Staden, R., (1982) Nucl. Acids Res. 10, 2951-2961.
13. Dayhoff, M.O., Barker, W.C., and Hunt, L.T. (1983) Meth. in Enzymol. 91, 524-545.
14. Barker, W.C. and Dayhoff, M.O. (1982) Proc. Natl. Acad. Sci. USA 79, 2836-2839.
15. Yulikova, E.P., Evseenko, L.K., Baratova, L.A., Belyanova, L.P. Rybin, V.K., and Silaev, A.B. (1976) Bioorg. Khim. 2, 1613-1617.
16. Yulikova, E.P., Rybin, V.K., and Silaev, A.B. (1979) Bioorg. Khim. 5, 5-10.
17. Ando, T. and Watanabe, S. (1969) Int. J. of Protein Res. 1, 221-224.
18. Iwai, K. Nakahara, C., and Ando, T. (1970) J. Biochem. 69, 493-509.
19. Debuire, B., Henry, C., Benaissa, M., Biserte, G., Claverie, J.M., Saule, S., Martin, P., and Stehelin, D. (1984) Science 224, 1456-1459.
20. Sternberg, M.J.E. and Taylor, W.R. (1984) FEBS Lett. 175, 387-392.
21. Möller, W. and Amons, R. (1985) FEBS Lett. 186, 1-8.
22. Weinmaster, G., Zoller, M.J., and Pawson, T. (1986) EMBO J. 5, 69-76.
23. Migliaccio, A., Rotondi, A., and Auricchio, F. (1984) Proc. Natl. Acad. Sci. USA 81, 5921-5925.
24. Baldi, A., Boyle, D.M., and Wittliff, J.L. (1986) Biochem. Biophys. Res. Comm. 135, 597-606.