Lemma 1.21. For every grammar, there is a variable-based grammar that is equivalent to it.

Proof. Add new variables X_a for every letter $a \in \Sigma$; let $V' := V \cup \{X_a\,;\, a \in \Sigma\}$. For each production rule $\alpha \to \beta \in P$, replace every occurrence of a letter a occurring in α by the corresponding new variable X_a; we write $X(\alpha)$ for this string. Clearly, $X(\alpha)$ does not contain any letters anymore, and so $X(\alpha) \to X(\beta)$ is a variable-based rule. Now define $P' := \{X(\alpha) \to X(\beta)\,;\, \alpha \to \beta \in P\} \cup \{X_a \to a\,;\, a \in \Sigma\}$ and $G' := (\Sigma, V', P', S)$.

Any G-derivation is transformed to a G'-derivation by the operation $\alpha \mapsto X(\alpha)$; a G-derivation of w becomes a G'-derivation of $X(w)$. Similarly, if we have a G'-derivation that contains no letters anywhere, then all strings occurring are of the form $X(\alpha)$ for some $\alpha \in \Omega^*$ and the operation of replacing all occurrences of X_a with the corresponding a transforms that derivation into a G-derivation. Together, this shows that $w \in \mathcal{L}(G)$ if and only of $X(w) \in \mathcal{D}(G', S)$.

If $X(w) \in \mathcal{D}(G', S)$, then (by applying the additional rules of the form $X_a \to a$ as needed) we have $w \in \mathcal{L}(G')$.

Conversely, assume that $w \in \mathcal{L}(G')$ and let $S = \sigma_0 \xrightarrow{G'}_1 \dots \xrightarrow{G'}_1 \sigma_m = w$ be a G'-derivation of w. If we apply the operation X to this derivation, we obtain a sequence $(\tau_0, \dots, \tau_m)$ with $\tau_0 = S = \sigma_0 = X(\sigma_0)$ and $\tau_i = X(\sigma_i)$. This sequence is not necessarily a G'-derivation. If $\sigma_i \xrightarrow{G'}_1 \sigma_{i+1}$ was an application of a rule of the form $X(\alpha) \to X(\beta)$, then the same rule will warrant that $X(\sigma_i) \xrightarrow{G'}_1 X(\sigma_{i+1})$; if $\sigma_i \xrightarrow{G'}_1 \sigma_{i+1}$ was an application of one of the rules $X_a \to a$, then applying X will result in $X(\sigma_i) = X(\sigma_{i+1})$. Since for each letter a there is only one production rule that produces a, we know that $|w|$ many steps of the derivation must be of this form. Thus, removing these $|w|$ many steps will make the remainder of the sequence $(\tau_0, \dots, \tau_m)$ a G'-derivation of length $m - |v|$ of $X(w)$. But then $w \in \mathcal{L}(G)$ by our earlier observation. Q.E.D.